

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИМЕНИ АКАДЕМИКА С. П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

На правах рукописи

Фролов Олег Олегович

**РАЗРАБОТКА СПОСОБА АНАЛИЗА СПЕКТРОВ КОМБИНАЦИОННОГО
РАССЕЯНИЯ ДЛЯ ПРИМЕНЕНИЯ В СТОМАТОЛОГИИ**

2.2.12. Приборы, системы и изделия медицинского назначения

Диссертация на соискание учёной степени
кандидата технических наук

Научный руководитель:

Тимченко Павел Евгеньевич,

кандидат физико-математических наук, доцент

Самара – 2025

Оглавление

ВВЕДЕНИЕ.....	5
1 ОПТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА БИОМАТЕРИАЛОВ В СТОМАТОЛОГИИ	16
1.1 Методы анализа дентинных материалов при деминерализации.....	16
1.2 Методы анализа брефоматрикса и ювенильного дентина	21
1.3 Методы диагностики пародонтита и оценки его лечения	22
1.3.1 Описание проблематики	22
1.3.2 Физические методы	25
1.3.3 Другие методы.....	29
1.3.4 Спектроскопия комбинационного рассеяния	32
1.4 Выводы по главе.....	36
2 ОБЗОР МЕТОДОВ ОБРАБОТКИ СПЕКТРАЛЬНЫХ ДАННЫХ КОМБИНАЦИОННОГО РАССЕЙЯНИЯ МНОГОКОМПОНЕНТНЫХ СРЕД	38
2.1 Спектроскопия комбинационного рассеяния и общая схема способа обработки данных	38
2.1.1 Интерполяция	39
2.1.2 Удаление случайных высокоинтенсивных шумов	40
2.1.3 Конвертирование единиц измерения длины волны	40
2.1.4 Нормализация.....	40
2.1.5 Сглаживание	42
2.2 Алгоритмы коррекции базовой линии спектров КР.....	45
2.2.1 Алгоритм многополиномиального аппроксимирования ModPoly	47
2.2.2 Алгоритм многополиномиального аппроксимирования I-ModPoly	51
2.2.3 Другие алгоритмы коррекции базовой линии	52
2.3 Обзор алгоритмов декомпозиции спектров КР.....	53
2.4 Выводы по главе.....	58
3 РАЗРАБОТКА АЛГОРИТМА ДЕКОМПОЗИЦИИ СПЕКТРОВ КОМБИНАЦИОННОГО РАССЕЙЯНИЯ БИОМАТЕРИАЛОВ	59
3.1 Общая схема алгоритма.....	60

3.2	Описание алгоритма итеративного поиска состава линий	62
3.3	Анализ моделей после итерационного поиска состава линий	69
3.4	Оценка влияния шума на качество декомпозиции	77
3.5	Оценка неопределенности.....	82
3.6	Выводы по главе.....	84
4	ПРИМЕНЕНИЕ РАЗРАБОТАННОГО СПОСОБА АНАЛИЗА СПЕКТРОВ КОМБИНАЦИОННОГО РАССЕЯНИЯ ДЛЯ ОЦЕНКИ ДЕНТИННЫХ МАТЕРИАЛОВ	86
4.1	Применение способа спектроскопии КР для оценки дентинных материалов в процессе деминерализации	86
4.1.1	Пробоподготовка и алгоритмы обработки спектров КР.....	86
4.1.2	Анализ усредненных спектров КР	88
4.1.3	Отбор признаков	89
4.1.4	Статистический анализ различий между группами	90
4.1.5	Процесс создания модели машинного обучения.....	93
4.1.6	Оценка качества и интерпретация полученного классификатора	99
4.2	Сравнительная характеристика биоматериалов из ювенильного дентина и брефоматрикса	107
4.2.1	Пробоподготовка и алгоритмы обработки спектров КР.....	107
4.2.2	Анализ усредненных спектров КР	109
4.2.3	Статистический анализ различий между группами	111
4.3	Выводы по главе.....	113
5	ЭКСПРЕСС-ДИАГНОСТИКА ПАРОДОНТИТА И ОЦЕНКА ЕГО ЛЕЧЕНИЯ С ПОМОЩЬЮ РАЗРАБОТАННОГО СПОСОБА	116
5.1	Оптическая диагностика изменений эмали зубов при пародонтите	116
5.1.1	Пробоподготовка и алгоритмы обработки спектров КР.....	116
5.1.2	Анализ усредненных спектров КР и результаты декомпозиции	118
5.1.3	Отбор признаков	121
5.1.4	Статистический анализ различий между группами	123
5.1.5	Процесс создания модели машинного обучения.....	126

5.1.6 Оценка качества и интерпретация полученного классификатора	128
5.1.7 Оценка качества модели без использования декомпозиции	133
5.1.8 Анализ <i>in vivo</i> спектров КР эмали для валидации полученной модели	135
5.2 Применение спектроскопии КР для оценки состояния костной ткани при пародонтите.....	137
5.2.1 Пробоподготовка и алгоритмы обработки спектров КР.....	137
5.2.2 Анализ усредненных спектров КР и результаты декомпозиции	138
5.2.3 Отбор признаков	140
5.2.4 Статистический анализ различий между группами	142
5.2.5 Процесс создания модели машинного обучения.....	144
5.2.6 Оценка качества и интерпретация полученного классификатора	146
5.3 Применение спектроскопии КР для оценки твердых тканей зубов.....	151
5.3.1 Пробоподготовка и алгоритмы обработки спектров КР.....	152
5.3.2 Анализ усредненных спектров КР	154
5.3.3 Статистический анализ различий между группами	156
5.4 Выводы по главе.....	159
ЗАКЛЮЧЕНИЕ	163
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ	165
СЛОВАРЬ ТЕРМИНОВ.....	167
СПИСОК ЛИТЕРАТУРЫ.....	172
ПРИЛОЖЕНИЕ А. Акт о внедрении результатов диссертационной работы в СамГМУ	186
ПРИЛОЖЕНИЕ Б. Акт о внедрении результатов диссертационной работы в «Центр восстановительной стоматологии».....	187

ВВЕДЕНИЕ

Диссертация посвящена разработке способа анализа спектров комбинационного рассеяния для диагностики стоматологических заболеваний и оценки состава биоматериалов, с целью повышения точности, достоверности и информативности спектроскопических исследований в клинической стоматологии и биомедицине.

Актуальность темы.

Высокая распространенность заболеваний пародонта, достигающая, по данным некоторых авторов, 98 % среди взрослого населения, подразумевает актуальность исследований, направленных на разработку методов диагностики пародонтита [1] [2] [3]. На сегодняшний момент в большинстве случаев заболевание выявляют лишь на стадии активных клинических проявлений. Часто диагностика осложняется тем, что к пародонтиту приводит множество факторов. Нередко он является следствием других заболеваний [4]. Как и при проблемах любой нозологии, ранняя диагностика пародонтита, вероятно, позволила бы купировать активный процесс и значительно улучшить прогноз заболевания. На сегодняшний день золотым стандартом диагностики пародонтита являются рентгеновские методы исследования, но они не всегда способны выявить ранние признаки воспалительного процесса, поскольку костные изменения наблюдаются лишь на более поздних стадиях заболевания. Актуальной задачей является развитие средств и методов диагностики стоматологических заболеваний, позволяющих врачу произвести своевременную и точную диагностику.

Степень разработанности темы.

В современном состоянии развития диагностики пародонтита применяются различные методы, каждый из которых обладает своими сильными и слабыми сторонами. Так, рентгеновские методы, являющиеся золотым стандартом в диагностике, демонстрируют высокую объективность, контрастность и детализацию костных структур, что позволяет достигать точности порядка 0,91 (0,83 – 0,96, 95 % доверительный интервал (ДИ)) с чувствительностью до 0,90

(0,82 – 0,96, 95 % ДИ) (J.Jarupat [5]). Однако, они не всегда выявляют ранние признаки воспалительного процесса, используют ионизирующее излучение и зависят от опыта оператора.

Флюоресцентная спектроскопия характеризуется высокой чувствительностью к изменениям минерального состава и доступностью оборудования, однако ее молекулярная специфичность ограничена из-за широкого спектра перекрывающихся флюоресцентных сигналов (Y.Park [6]).

При этом ИК-спектроскопия, применяемая для анализа белков, липидов и углеводов, демонстрирует точность классификации до 0,95 в сочетании с методами машинного обучения (S.Fujii [7], S.Silva [8]), но ограничена из-за чувствительности к присутствию воды и недостаточной пространственной разрешающей способности.

Спектроскопия комбинационного рассеяния (КР) ранее использовалась с целью различения трёх основных пародонтопатогенов Авторами (J.Zhang и др. [9]) заявлена точность классификации на уровне ROC AUC 0,94.

Особое внимание уделено методам анализа дентинных материалов при деминерализации. Исследования в этой области активно используют оптическую когерентную томографию (ОКТ), в том числе swept-source ОКТ, как показано в работе Matsuzaki (2021) [10], а также поляризационно-чувствительную оптическую когерентную томографию (ОКТ) (PS-ОКТ), продемонстрированную в исследовании Manesh (2009) [11] для неразрушающей оценки кариозных изменений в дентине. Дополнительно, исследование Koga (2016) [11] показало, что частичная деминерализация дентина способствует более выраженной регенерации костной ткани, что подтверждает важность выбора оптимальных методов оценки. Кроме того, для анализа процессов деминерализации используются различные спектроскопические методы, позволяющие фиксировать изменения в химическом составе дентина, что дополняется вспомогательными методами, такими как ручная деконволюция спектров и анализ ультраструктурных характеристик с использованием РЭМ. Также применяется и ультразвуковая диагностика (Denis и

др. 2018 [12]) для классификации уровней деминерализации костной ткани. Максимальная точность и специфичность классификации достигала 0,80 и 0,83.

Учитывая как возможности методов, используемых для диагностики пародонтита, так и особенности методик анализа дентинных материалов при деминерализации, спектроскопия комбинационного рассеяния представляется перспективным инструментом, способным объединить преимущества различных подходов. Спектроскопия КР неинвазивна, обеспечивает высокую молекулярную детализацию, позволяет проводить исследования *in vivo* без разрушительной подготовки образцов и эффективно работает в водной среде, что делает её незаменимым элементом в динамическом мониторинге состояния зубных тканей. При этом метод сталкивается с проблемами влияния флюоресцентного фона, сложностями алгоритмической обработки и высокой требовательностью к калибровке оборудования.

Именно поэтому данная диссертационная работа направлена на разработку и оптимизацию способа анализа спектров комбинационного рассеяния, что позволит решить указанные проблемы и объединить достоинства различных способов для повышения точности и надежности диагностики пародонтита.

Целью диссертационного исследования является разработка и экспериментальная валидация способа анализа спектров комбинационного рассеяния биоматериалов, позволяющего проводить неинвазивный контроль состояния твердых тканей зубов и костной ткани при пародонтите, а также оценку биоматериалов.

Задачи диссертационного исследования.

1. Разработка способа анализа спектров комбинационного рассеяния биоматериалов, включающего алгоритмы предобработки и декомпозиции.

2. Определение возможности оценки состояния дентинных материалов по спектральным критериям, для оценки их деминерализации на основе выделения ключевых амплитудных признаков, с помощью спектроскопии комбинационного рассеяния.

3. Апробация способа экспресс-диагностики пародонтита на основе анализа спектров комбинационного рассеяния эмали и костной ткани с целью оценки его диагностической эффективности.

Соответствие паспорту специальности 2.2.12 Приборы, системы и изделия медицинского назначения.

Область исследования соответствует следующим пунктам паспорта специальности 2.2.12. Приборы, системы и изделия медицинского назначения (отрасль науки – технические):

14. Методы, модели и алгоритмы, включая распознавание образов, для медицинских информационных и интеллектуальных систем обеспечивающих повышение эффективности медико-биологических исследований и врачебных решений;

21. Методы и средства искусственного интеллекта для медико-биологических исследований;

22. Специальное программное обеспечение, предназначенное для профилактики, диагностики, лечения и медицинской реабилитации заболеваний, мониторинга состояния организма человека и проведения медицинских исследований.

Научная новизна результатов исследования:

1. Разработан итеративный алгоритм декомпозиции спектров комбинационного рассеяния биологических материалов, основанный на автоматической оптимизации числа линий, их амплитуд, положений и ширин. Алгоритм при взаимном перекрытии полос до 75 % обеспечивает точность определения позиций линий не хуже $0,1 \text{ см}^{-1}$, относительную погрешность амплитуд и ширин $\leq 0,3 \%$ при детерминации модели $R^2 \geq 0,995$.

2. Предложен способ оценки деминерализации дентинных материалов, основанный на извлечении информативных признаков из декомпозированных спектральных линий комбинационного рассеяния в диапазоне $630\text{--}1450 \text{ см}^{-1}$. Полученные признаки используются в модели логистической регрессии для вероятностной классификации образцов по наличию признаков деминерализации с

точностью ROC AUC 1,0 (0,99–1,0, 95 % ДИ) для определения факта деминерализации.

3. Предложен способ экспресс-диагностики пародонтита, основанный на выявлении структурных признаков изменения ткани зуба при комплексной обработке декомпозированных спектральных данных комбинационного рассеяния с применением логистической регрессии. Способ позволяет диагностировать признаки пародонтита при анализе спектров комбинационного рассеяния костной ткани на уровне ROC AUC 0,88 (0,78-0,96, 0,95 ДИ), а для спектров эмали ROC AUC составляет 0,94 (0,89-0,98, 0,95 ДИ) при чувствительности и специфичности классификации пародонтита на уровне 0,86 (0,76-0,94, 0,95 ДИ) и 0,86 (0,76-0,94, 0,95 ДИ).

Основные положения, выносимые на защиту:

1. Применение итеративного алгоритма декомпозиции спектров комбинационного рассеяния биологических материалов позволяет обеспечить точность определения положения центров спектральных линий не хуже $0,1 \text{ см}^{-1}$ с относительной погрешностью определения амплитуды и ширины линии ниже 0,3 % при взаимном перекрытии полос до 75 % и детерминации модели R^2 не ниже 0,995. (Соответствие пункту 22 паспорта специальности).

2. Использование предложенного способа оценки деминерализации дентинных материалов по спектрам комбинационного рассеяния, с применением логистической регрессии для вероятностной классификации образцов по информативным признакам декомпозированных спектральных линий, позволяет проводить диагностику с точностью на уровне ROC AUC 1,0 (0,99-1,0, 0,95 ДИ). (Соответствие пункту 14, 21 паспорта специальности).

3. Использование предложенного способа экспресс-диагностики пародонтита на основе спектров комбинационного рассеяния эмали и костной ткани, обладающего чувствительностью к ранним изменениям минерального состава, с применением логистической регрессии, позволяет диагностировать признаки пародонтита на уровне ROC AUC 0,94 (0,89-0,98, 0,95 ДИ) для эмали и

0,88 (0,78-0,96, 0,95 ДИ) для костной ткани. (Соответствие пункту 14, 21 паспорта специальности).

Теоретическая значимость работы заключается в разработке и обосновании нового подхода к анализу спектров КР с использованием декомпозиции спектров комбинационного рассеяния с высоким коэффициентом перекрытия элементарных линий (75 %), что позволяет проводить количественный анализ сложных спектров биоматериалов. Предложенный алгоритм учитывает вариативность ширины и положения спектральных линий, а также устойчив к флюоресцентному фону и шумам, что расширяет применимость метода КР в биомедицинских исследованиях.

Разработанная модель обработки спектров, включающая предобработку, декомпозицию и анализ признаков, может быть положена в основу новых теоретических подходов к интерпретации спектральных данных в задачах диагностики и контроля состава биотканей. Работа также формирует основу для построения интеллектуальных систем анализа биомедицинских спектров с применением методов машинного обучения.

Практическая значимость заключается в разработке программного обеспечения, реализующего разработанный способ анализа спектров комбинационного рассеяния для диагностики пародонтита и оценки качества биоимплантов в процессе их производства.

Способ позволяет выявлять признаки пародонтита, в том числе на ранних стадиях развития заболевания, оценивать деминерализацию дентинных материалов с промежуточными значениями.

Результаты диссертационного исследования внедрены в научно-исследовательском институте биотехнологий Самарского государственного медицинского университета (НИИ БиоТех СамГМУ) (ПРИЛОЖЕНИЕ А) (акт о внедрении результатов кандидатской диссертационной работы № 1230 / 40-23-3345 от 10.07.2025) и в стоматологической клинике «Центр восстановительной стоматологии» (ПРИЛОЖЕНИЕ Б) (акт о внедрении результатов кандидатской диссертационной работы от 15.01.2025).

Результаты, вошедшие в диссертационное исследование, были использованы при выполнении работ по теме: «Разработка программного обеспечения для идентификации и аттестации компонентов клеточно-тканевых имплантатов с помощью оптического метода при их производстве» (договор № 14483ГУ/2019 от 17.07.2019 г. с Фондом содействия инновациям). А также были использованы при выполнении работ по проекту 18-315-20017 мол_а_вед «Новые фундаментальные подходы в комплексной оценке структуры костной ткани в норме, при экспериментальном остеопорозе и его профилактики с помощью продуктов биотехнологий».

Достоверность полученных результатов.

Достоверность результатов обеспечивается строгостью применённых методов обработки данных и машинного обучения, а также воспроизводимостью экспериментальных измерений. Корректность интерпретации спектров подтверждается статистическим анализом (ROC-AUC, *p*-значение) и соответствием полученных данных результатам инструментальных и клинических методов оценки. Основные результаты апробированы в профильных учреждениях и опубликованы в рецензируемых научных изданиях, входящих в базы данных Scopus, Web of Science и ВАК.

Методы исследования.

В работе применялись методы экспериментальной спектроскопии комбинационного рассеяния, а также методы цифровой обработки сигналов, математического моделирования, статистического анализа и машинного обучения. Спектральные данные получены с использованием рамановского спектрометра с лазерным источником возбуждения на длине волны 785 нм.

Первичная предобработка спектров включала: интерполяцию, преобразование длины волны в рамановский сдвиг [см^{-1}], нормализацию (по площади под кривой), а также сглаживание сигналов с использованием фильтра максимального правдоподобия MLE-SG на основе алгоритма Савицкого-Голея. Коррекция флуоресцентной составляющей спектров и базовой линии осуществлялась с использованием алгоритм Improved Modified Multi-Polynomial

Fitting (улучшенный модифицированный алгоритм многополиномиального аппроксимирования (I-ModPoly)).

Отличительной особенностью предложенного подхода является использование оригинального итеративного алгоритма декомпозиции спектров, разработанного автором, который учитывает наложение спектральных линий с коэффициентом перекрытия до 75 %.

Для построения моделей классификации использовались алгоритмы логистической регрессии, а также методы машинного обучения с контролем переобучения. Качество моделей оценивалось по метрикам точности, чувствительности, специфичности и площади под ROC-кривой (AUC).

Вся математическая обработка и визуализация данных проводилась с использованием языка программирования Python в средах разработки Jupyter Notebook и PyCharm. Использовались библиотеки NumPy, SciPy, Pandas, Matplotlib, scikit-learn и другие специализированные модули для обработки спектроскопических данных.

Личный вклад автора.

Личный вклад автора заключается в самостоятельной разработке способа обработки и анализа спектров комбинационного рассеяния, включая создание оригинального алгоритма декомпозиции спектров с высоким коэффициентом перекрытия линий. Автором лично выполнено программное обеспечение, реализующее предложенные алгоритмы, включая автоматизированную обработку спектральных данных и построение моделей машинного обучения для диагностики стоматологических заболеваний. Также автором проведена апробация разработанных способов на реальных биологических образцах, статистический анализ, визуализация и интерпретация полученных результатов.

Все оригинальные результаты, представленные в диссертации, получены автором лично либо при его непосредственном участии. Постановка задач и обсуждение результатов осуществлялись совместно с научным руководителем.

Публикации.

По теме диссертации опубликовано 7 статей в научных журналах и изданиях, рекомендованных Высшей аттестационной комиссией и/или рецензируемых базой данных Scopus и WoS. Получено 1 свидетельство о регистрации РИД (патент на изобретение: «Способ одновременного получения деминерализованного дентина и минеральноорганического компонента из зубов» RU 2 752 035С1).

Наиболее значимые работы по теме диссертации:

1. Decomposition Method for Raman Spectra of Dentine / О. О. Frolov и др. // J of Biomedical Photonics & Eng. – 2024. – Vol. 10, No. 3. – 030303-1 – DOI 10.18287/JBRE24.10.030303.
2. Применение метода спектроскопии комбинационного рассеяния для оценки дентинных материалов в процессе их изготовления / П. Е. Тимченко, Е. В. Тимченко, Л. Т. Волова, О. О. Фролов // Оптический журнал. – 2021. – Т. 88, № 9. – С. 3-8. – DOI 10.17586/1023-5086-2021-88-09-03-08.
3. Хемометрический анализ биоимплантатов из твердой мозговой оболочки при их изготовлении / П. Е. Тимченко, Е. В. Тимченко, Л. Т. Волова, О.О. Фролов // Оптический журнал. – 2019. – Т. 86, № 1. – С. 13-20. – DOI 10.17586/1023-5086-2019-86-01-13-20.
4. Экспериментальные исследования гидроксиапатита методом спектроскопии комбинационного рассеяния / П. Е. Тимченко, Е. В. Тимченко, Е. В. Писарева, М.Ю. Власов, Л.Т. Волова, О.О. Фролов, А.Р. Калимуллина // Оптический журнал. – 2018. – Т. 85, № 3. – С. 12-18.
5. Optical Assessment of In-Office Teeth Whitening Influence on Tooth Enamel Spectral Composition In Vivo / E. V. Timchenko, O. O. Frolov, O. A. Magsumova [и др.] // Optical Memory and Neural Networks. – 2022. – Vol. 31, No. 4. – P. 413-417. – DOI 10.3103/s1060992x22040099.
6. Raman spectroscopy for assessment of hard dental tissues in periodontitis treatment / E. V. Timchenko, O. O. Frolov, P. E. Timchenko [и др.] // Diagnostics. – 2021. – Vol. 11, №. 9. – DOI 10.3390/diagnostics11091595.

7. Raman spectroscopy of changes in the tissues of teeth with periodontitis / E. V. Timchenko, O. O. Frolov, P. E. Timchenko [и др.] // *Diagnostics*. 2020. – Vol. 10. № 9. – DOI 10.3390/diagnostics10110876.

Зарегистрированные результаты интеллектуальной деятельности:

1. Патент № 2752035 С1 Российская Федерация, МПК А61К 35/32, А61К 6/80, А61L 27/02. Способ одновременного получения деминерализованного дентина и минерально-органического компонента из зубов: № 2021100652: заявл. 13.01.2021: опубл. 22.07.2021 / Л. Т. Волова, О. О. Фролов, П. Е. Тимченко [и др.].

Апробация результатов диссертации.

Результаты диссертационной работы были представлены на:

X международной конференции по информационным технологиям и нанотехнологиям (ITNT-2024), г. Самара (2024 г.);

XXI Всероссийском молодежном Самарском конкурсе-конференция по оптике, лазерной физике и физике плазмы, г. Самара (2022, 2023 гг.);

Международной конференции для молодых ученых и студентов Saratov Fall Meeting – International School for Junior Scientists and Students on Optics, Laser Physics & Biophotonics, г. Саратов (2022, 2023 гг.);

Международной конференция ALT`23 (2023 г.), г. Самара; Всероссийской научной конференции с международным участием «Енисейская Фотоника» (2022 г.), г. Красноярск;

Всероссийской научно-практической конференции с международным участием «Аспирантские чтения» (2022, 2023 гг.), г. Самара;

Международной конференции «Лазерно-информационные технологии» (2021, 2022 гг.), г. Новороссийск;

Международной конференции «Fundamentals of laser-assisted micro- and nanotechnologies» (FLAMN-22) (2022 г.), г. Санкт-Петербург;

Международной конференции «Физика.СПб» (2021 г.), г. Санкт-Петербург;

Международной конференции «EEhPolytech» (2021 г.), г. Санкт-Петербург;

Международной конференции «European Conferences on Biomedical Optics Conference» (2021 г.), г. Мюнхен.

Структура и объем диссертации.

Диссертация состоит из введения, пяти глав, заключения, списка сокращений и условных обозначений, списка литературы, включающего 149 наименований, и двух приложений. Работа изложена на 187 страницах машинописного текста, содержит 76 рисунков, 16 таблиц.

1 ОПТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА БИОМАТЕРИАЛОВ В СТОМАТОЛОГИИ

В данной главе рассматриваются современные оптические и спектроскопические методы анализа биоматериалов в стоматологии, с особым акцентом на диагностику заболеваний пародонта и оценку их лечения. Глава демонстрирует, как неинвазивные методы визуализации и анализа, включая различные спектроскопические техники, позволяют получить детальную информацию о микроструктурных и химических изменениях в тканях зубов и костей, что имеет решающее значение для ранней диагностики, мониторинга и коррекции патологических процессов.

Основная цель главы – обобщить современные подходы и способы, применяемые для неразрушающего анализа дентина, эмали и костных структур, а также выявить преимущества и ограничения каждого метода. Особое внимание уделено сравнению традиционных методов диагностики пародонтита с современными спектроскопическими техниками, что позволяет показать потенциал развития более точных, быстрых и безопасных способов мониторинга состояния зубочелюстной системы.

1.1 Методы анализа дентинных материалов при деминерализации

Необходимость костной пластики в хирургической стоматологии необходима в первую очередь пациентам с отсутствующими зубами. Частичное или полное отсутствие зубов всегда сопровождается признаками атрофии костной ткани челюстей, что применительно к использованию дентальных имплантатов является показанием к устранению дефицита объема альвеолярной кости не менее чем в 30 % клинических случаев. В таких ситуациях требуется либо предварительное восстановление костной ткани, либо одномоментная реконструкция с установкой имплантатов [13].

В последнее время на медицинском рынке наблюдается большое разнообразие костнопластических материалов, имеющих как биологическое

происхождение, так и синтезированных искусственных медицинских изделий. Среди них широкое распространение получили следующие типы материалов: биологические костно-пластические материалы, аутотрансплантаты, костные аллоимплантаты, ксеноимплантаты и синтетические костно-пластические материалы [14]. Однако применяемые в настоящее время материалы не лишены недостатков, к которым относятся их высокая стоимость, большая степень усадки материала, частичное отторжение материала при повторном вскрытии операционного поля для углубления имплантатов и др.

Поиск новых материалов в данной области непрерывно продолжается [15]. Одним из кандидатов на роль такого материала является человеческий дентин [16].

В работе [16] показано, что при изготовлении биоматериалов из дентина его необходимо подвергать процедуре декальцинирования для лучшей его приживляемости. Однако для увеличения регенеративных функций биоматериалов из дентина требуется сохранить в процессе деминерализации органические компоненты, необходимые для костеобразования.

Для оценки биоматериалов из дентина в процессе их изготовления и подбора оптимальных щадящих методов обработки целесообразно использовать неdestructивные методы анализа.

Исследования по оценке деминерализации дентина зубов демонстрируют активное развитие методик, позволяющих не только оценивать утрату минералов, но и контролировать эффективность различных лечебных средств. В работах использованы методы оптической когерентной томографии (ОКТ), РЭМ, спектроскопии и микротомографии, что позволяет получить глубокое понимание процессов де- и реминерализации.

В работе Matsuzaki 2021 [10] исследовались механизмы действия и ингибирующее действие двух дезсенсibilizаторов (Nanoseal и Caredyne Shield) на деминерализацию дентина. Использовалась swept-source оптическая когерентная томография (SS-OCT) для получения двумерных изображений образцов, а также сканирующая электронная микроскопия и энергодисперсионная рентгеновская спектроскопия для анализа морфологии и химического состава. Результаты

показали, что оба материала обладают эффектом ингибирования дальнейшей деминерализации, что подтверждается изменениями в показателях интенсивности и коэффициента затухания сигнала ОКТ.

В статье автора Manesh, 2009 [11] проведено исследование возможности использования поляризационно-чувствительной оптической когерентной томографии (PS-ОКТ) для неразрушающей оценки кариозных изменений в дентинах. Авторы продемонстрировали, что PS-ОКТ может эффективно измерять глубину деминерализации и оценивать степень ингибирования деминерализации после применения профилактических агентов, таких как фторид и лазерное воздействие.

Исследование Koga 2016 [11] посвящено влиянию деминерализации и размера частиц матрицы дентина на регенерацию костной ткани. Результаты показали, что частично деминерализованная матрица дентина (PDDM) с крупными частицами (около 1000 мкм) обеспечивает более выраженную регенерацию костной ткани, вероятно, за счёт оптимальной поверхности для адгезии клеток и сбалансированного процесса резорбции и формирования новой костной ткани.

В исследовании Denis и др. 2018 [12] применялась ультразвуковая диагностика в сочетании с алгоритмом SVM для классификации уровней деминерализации костной ткани. Были выделены акустические признаки, такие как скорость распространения волн и среднеквадратичное значение сигнала (RMS). Максимальная чувствительность и специфичность классификации между интактной костью и костью, деминерализованной в течение 72 часов, достигала 80 % и 82,61 % при использовании. Для других уровней деминерализации точность также была на уровне 80 % при различных объемах обучающей выборки.

Usenik и др. (2015) [17] продемонстрировали возможность автоматизированной классификации и визуализации здоровых и патологических (включая очаги деминерализации) зон твёрдых тканей зуба с помощью ближней инфракрасной (NIR) гиперспектральной съёмки и многочленной дискриминантной анализа. Для 12 образцов зубов они получили чувствительность 85 % и специфичность 97 % при разграничении здоровой и поражённой ткани.

Эти публикации демонстрируют, что тема оценки деминерализации дентина зубов активно исследуется с использованием современных методов визуализации и анализа. Разработаны как экспериментальные модели деминерализации, так и методы неинвазивного контроля процессов реминерализации, что предоставляет широкие возможности для дальнейших клинических исследований и разработки новых профилактических и лечебных стратегий.

Спектроскопия комбинационного рассеяния является перспективным методом для неразрушающей оценки деминерализации дентина. Метод позволяет получать детальную информацию о химическом составе и структурных изменениях в зубной ткани, что особенно ценно для ранней диагностики и мониторинга эффективности профилактических или лечебных мер. Спектроскопия КР также имеет то преимущество, что она неинвазивна и масштабируема для портативных устройств, сохраняя при этом хороший уровень чувствительности.

Tramini и др. [18] исследовали *in vitro* влияние молочной кислоты на эмаль и дентин постоянных зубов человека. Результаты выявили изменения, вызванные кислотным воздействием, главным образом в отношении фосфата PO_4^{3-} , который представляет собой минеральную фазу в эмали и дентине. Процесс восстановления минералов внутри дентина и эмали оценивали путем создания *in vitro* низкого pH, использованные образцы из срезов зубов погружали в чистый раствор молочной кислоты (5M, pH = 4,5) на 7, 14 и 30 дней. После этого образцы промывали дистиллированной водой перед анализом. Спектры комбинационного рассеяния показали изменения интенсивности симметричной полосы растяжения PO_4^{3-} , которые зависят от типа дентина, анатомического расположения и возраста субъекта.

В статье авторов [19] описано использование рамановской спектроскопии для анализа структуры зубов по сравнению с синтетическими апатитами. Обсуждаются спектральные линии, связанные с твердыми тканями зуба, которые предоставляют важные данные для понимания химических структурных свойств дентина и эмали. Также была использована ручная деконволюция спектров с помощью метода наименьших квадратов.

В работе [20] с помощью спектроскопии КР показано, что при реминерализации химический состав минеральной фазы можно было легко восстановить, в то время как потеря объема минералов проявлялась как почти необратимый эффект, особенно для дентина, подвергающегося воздействию низких значений pH.

В работе Marin 2020 [21] исследователи систематически мониторили процессы деминерализации и реминерализации дентина с использованием рамановской спектроскопии. Они продемонстрировали, что два параметра рамановского спектра – ширина полосы симметричного растяжения PO_4 (960 см^{-1}) и отношение интенсивностей минерального и органического компонентов – коррелируют с изменениями структуры дентина при деминерализации, что позволяет использовать рамановскую спектроскопию для мониторинга этих процессов в реальном времени.

Исследование Zhan 2013 [22] посвящено оценке изменений в составе человеческого дентина при деминерализации и реминерализации с использованием рамановской спектроскопии. Авторы показали, что интенсивности пиков, соответствующих минеральным компонентам (например, PO_4), снижаются с увеличением времени экспозиции к кислоте, а после реминерализации интенсивности этих пиков увеличиваются. Таким образом, спектроскопия КР может служить эффективным инструментом для оценки де- и реминерализации дентина.

Также используется витальное окрашивание - применение красителей, таких как метиленовый синий, для выявления деминерализованных участков. Интенсивность окрашивания отражает степень деминерализации. Этот метод прост в применении, но менее точен по сравнению с аппаратными методами.

Другим более субъективным методом оценки является проверка пластичности подручными медицинскими инструментами, что также не позволяет точно определить степень деминерализации.

1.2 Методы анализа брeфоматриксoв и ювенильного дeнтина

При применении костных аллотрансплантатов высока вероятность иммуногенного ответа, существуют риски передачи инфекций, что приводит к ограниченности их использования. Концентрация примесей минеральной составляющей костной ткани оказывает значительное влияние на биологические, механические и оптические свойства [23].

При их применении, в отличие от ауто- и ксенопластики и использования синтетических препаратов, не нарушается гомеостаз и метаболизм соединительных тканей и функции систем жизнеобеспечения реципиента [24]. Аллогенные материалы после их специальной обработки практически полностью теряют свою антигенность и при помещении в организм не оказывают на него негативного воздействия. Они играют роль матрицы, кондуктора, постепенно полностью рассасываются, а на их месте формируется новая костная ткань [25].

Расширение источников получения аллогенных материалов является чрезвычайно важной и актуальной задачей.

Низкая антигенность и высокие остеоиндуктивные свойства обуславливают применение брeфокости в качестве пластического материала. Костный брeфоматрикс после деминерализации сохраняет эластичность, что позволяет облегчить процесс моделирования трансплантата [26].

Брeфоматрикс представляет собой деминерализованный костный матрикс, который получают путём обработки костной ткани – обычно это кости плодов человека на 18–22 недели гестации (иногда используют также ткани взрослых), – раствором соляной кислоты слабой концентрации. При таком методе удаляются минеральные компоненты, в то время как сохраняется органическая матрица, богатая коллагеном и биоактивными белками, что обеспечивает его остеогенную активность и способствует стимуляции регенеративных процессов в костной ткани.

Однако, в настоящее время, брeфоматриксy не изготавливаются в России, в связи с изменением Российского законодательства (Приказ Минздравсоцразвития РФ от 03.12.2007 N 736 (ред. от 27.12.2011) "Об утверждении перечня медицинских

показаний для искусственного прерывания беременности" (Зарегистрировано в Минюсте РФ 25.12.2007 N 10807)), согласно которому прерывание беременности не проводится на поздних сроках (18-24 недели) гестации плода, а проводится строго по медицинским показаниям.

Аналогом брeфоматриксу в качестве трансплантата, могут служить деминерализованные ювенильные зубы или брeфокость из утилизированных послеоперационных добавочных пальцев детей с полидактилией, содержащие большое количество структурных белков аналогичных костной ткани и брeфоматриксу.

Ювенильный дентин – это тип зубной ткани, который формируется в молочных (детских) зубах и характеризуется менее высокой степенью минерализации, более высоким содержанием органических компонентов (например, коллагена) и отличительной микроструктурой по сравнению с дентином постоянных зубов. Получают его преимущественно из молочных зубов, которые естественным образом теряются в детском возрасте или извлекаются по клиническим показаниям; затем такие образцы проходят этапы очистки, обеззараживания и подготовки для последующего исследования или применения в биомедицинских исследованиях.

Настоящее исследование представляет собой первое комплексное изучение спектральных характеристик брeфоматриксов и ювенильного дентина с использованием метода спектроскопии КР. По нашим данным, аналогичные работы в современной литературе отсутствуют, что подчеркивает новизну и оригинальность полученных результатов.

1.3 Методы диагностики пародонтита и оценки его лечения

1.3.1 Описание проблематики

Пародонтит – заболевание воспалительного характера, сопровождающееся деструктивным разрушением всех тканей пародонта [27]. Воспалительные заболевания пародонта (ВЗП) по-прежнему составляют актуальную, не до конца решенную проблему современной стоматологии в связи с сохраняющимся

высоким уровнем распространенности гингивита и пародонтита у населения различных популяций. В России распространенность признаков поражения пародонта в 12-15-летнем возрасте составляет 34-40 %, в возрастной группе 35-44 лет достигает 81 %, а у лиц старше 65 лет составляет 91 % [28] [29] [30].

Мягкий зубной налет (ЗН), вырабатывающий токсины (NH_3 , H_2S , лактат, индол и другие) и протеолитические ферменты, вызывает воспаление десны – гингивит. В результате накопления ЗН в течение 2-4 дней развивается воспалительный процесс (гингивит) в подэпителиальной соединительной ткани десневой бороздки. При его дальнейшем увеличении через 2-3 недели он проникает под десну, разрушая зубодесневое эпителиальное соединение.

По мере дальнейшего роста зубного налета воспалительные явления распространяются на костную ткань и соединительно-тканые волокна, удерживающие зуб в альвеолярной кости, может развиваться пародонтит [31].

В последние годы отмечается резкий рост заболеваемости пародонтитом. Заболевания пародонта являются медико-социальной проблемой, так как пародонтит вызывает потерю зубов, а инфекции пародонтальных карманов могут поражать организм в целом [32]. Воспалительные заболевания пародонта представляют собой одну из наиболее распространенных стоматологических проблем. Течение заболевания характеризуется нарастанием тяжести и интенсивности и формированием хронического пародонтогенного очага инфекции [33]. Хронический катаральный гингивит и хронический генерализованный пародонтит на ранних стадиях характеризуются маломанифестным или латентным периодом инфицирования [34].

Хронический пародонтит является одной из тяжелых и распространенных патологий пародонта, который приводит к значительному нарушению функций зубочелюстной системы, с повреждением удерживающего аппарата зубов и выпадению последних [1]. Наибольшей распространенности (60–65 %) пародонтит достигает у людей старше 30 лет [2]. При этом, доля пациентов с тяжелой формой хронического пародонтита увеличилась у молодых лиц до 11,2 %, и у людей в возрасте 65 лет и старше до 30 % [3].

Современные данные свидетельствуют, что пародонтит является полиэтиологическим заболеванием [35]. Коварство развития пародонтита заключается в том, что начальные признаки возникновения воспалительного процесса часто остаются незамеченными и хроническое течение заболевания, приводит к тяжелым последствиям не только для зубочелюстной системы, но и для организма пациента в целом.

Пародонтит является одним из факторов резорбции костной ткани зуба [36]. Так же пародонтит приводит к нарушению минерализации костных тканей зубов, разрушению органической матрицы с образованием резорбционных лакун [36] [37].

В большинстве работ по изучению изменений при этом заболевании основное внимание уделяется изучению не изменений твердых тканей зубов, а окружающих зуб мягких тканей [19] [38], ротовой жидкости [39] [40], а также восстановлению костной ткани при лечении пародонтита [41].

Лишь незначительная часть работ посвящена исследованию тканей зубов при пародонтите. Так в работе ряда авторов [42] показано, что основные структурные изменения в тканях зубов при генерализованном пародонтите проявляются в изменении минерализации дентина и одним из дополнительных способов исключения прогрессирования процессов деминерализации дентина был рекомендован процесс депульпирования таких зубов.

Авторами работ [43] отмечается повышение микротвердости эмали при прогрессировании пародонтита.

В тоже время авторы статьи [42] на основании проведенных исследований показали, что явных структурных изменений в эмали при генерализованном пародонтите не выявлено.

Анализ литературных данных показывает, что при хроническом пародонтите одним из дискуссионных вопросов является изменение твердых тканей зубов, в особенности эмали, т.к. имея информацию об изменении состава эмали зубов при пародонтите можно было бы разработать неинвазивный метод диагностики данного заболевания.

При лечении пациентов с заболеваниями пародонта является актуальной своевременная его диагностика и профилактика. Для решения данной задачи и неинвазивной диагностики пародонтита, необходимо знать какие структурные изменения происходят в тканях зубов при данном заболевании.

Поэтому, проведение исследований по изучению тканей зубов при пародонтите является весьма актуальной задачей.

1.3.2 Физические методы

Среди существующих методов исследования тканей зубов широкое распространение получили методы биохимического анализа, растровой электронной микроскопии [42], флюоресценции [39] и спектроскопии [42].

1.3.2.1 Рентгеновские методы

На сегодняшний день золотым стандартом диагностики пародонтита являются рентгеновские методы исследования, в частности традиционная рентгенография, а также её современные модификации, такие как цифровая панорамная рентгенография и компьютерная томография (КТ). В ряде работ рентгеновские методы демонстрируют высокую объективность в выявлении пародонтита. Например, интерпретация снимков компьютерной томографии (КТ) пародонтистом может достигать точности 0,92 (0,83–0,96, 0,95 ДИ), чувствительности до 0,91 (0,82–0,96, 0,95 ДИ) и специфичности 1,0 (0,48–1,0, 0,95 ДИ) при оценке степени потери костной ткани [5].

Преимущества данного подхода:

- Объективность, высокая контрастность и детализация. Позволяют получить чёткое изображение костных структур, оценить глубину пародонтальных карманов и объем потери костной ткани.
- Возможность последующего количественного анализа. Цифровая панорамная рентгенография легко интегрируется с системами для автоматизированного измерения параметров.

Недостатки метода:

- Рентгенография не всегда способна выявить ранние признаки воспалительного процесса, поскольку костные изменения наблюдаются лишь на более поздних стадиях заболевания.
- Применяется ионизирующее излучение, что ограничивает частоту повторных исследований.
- При диагностике мягких тканей и поверхностных изменений метод обладает ограниченной чувствительностью.
- Высокая стоимость оборудования. КТ-сканеры особенно дорогие и требуют специальных условий эксплуатации.
- Зависимость от навыков оператора. Правильная интерпретация изображений требует опыта и калибровки оборудования.

1.3.2.2 Растровая электронная микроскопия (РЭМ)

РЭМ обеспечивает качественные изображения микроструктуры тканей зубов и являются наиболее широко используемыми методами для оценки их структуры, но требуется деструктивная подготовка образца, что делает невозможным применение для динамического мониторинга в клинических условиях [44] [41] [42].

РЭМ позволяет визуализировать ультраструктурные особенности пародонтальных тканей с разрешением до нанометров. Точные числовые показатели точности как таковые не приводятся в клинических исследованиях, поскольку РЭМ используется преимущественно в исследовательских целях для изучения морфологии.

Преимущества:

- Высокое разрешение. Позволяет детально изучить микроструктуру тканей, выявить изменения на уровне клеток и внеклеточного матрикса.
- Объективная морфологическая оценка. Предоставляет детальные данные для научных исследований и понимания патогенеза.

Недостатки:

- Деструктивность подготовки образца. Обработка требует специального обезвоживания, покрытия металлом и работы в вакууме, что не допускает исследования живых образцов.
- Высокая стоимость и трудоемкость. Оборудование дорогостоящее, а проведение анализа требует значительных временных затрат и квалифицированного персонала.
- Ограниченная клиническая применимость. РЭМ больше подходит для фундаментальных исследований, чем для рутинной клинической диагностики.

1.3.2.3 Флюоресцентная спектроскопия

Флюоресцентная диагностика в стоматологии базируется на анализе спектров флюоресценции твердых тканей зуба. В качестве основного вещества при этом исследуется гидроксиапатит, из которого состоят зубы, при этом анализ детального состава зубов представляется невозможным. Метод демонстрирует высокую чувствительность, но ограничен в молекулярной специфичности [39].

Количественная светоиндуцированная флюоресценция (Quantitative Light-Induced Fluorescence QLF) используется для обнаружения кариеса путем получения флюоресцентных изображений зубов. Этот метод усовершенствован за счет машинного обучения, в частности использования сверточных нейронных сетей (CNN) для автоматического анализа и классификации изображений. Этот метод обеспечивает высокую чувствительность и специфичность, что делает его мощным инструментом для раннего обнаружения и мониторинга кариеса [6].

Хотя многие работы с QLF были изначально ориентированы на раннюю диагностику кариеса или оценку зубного налёта, ряд исследований демонстрирует, что QLF позволяет объективно оценивать состояние зубного налёта и его активность.

1.3.2.4 Метод оптического поляриметрического обнаружения (Optical Polarimetric Detection)

Метод использует свойства двойного лучепреломления зубной эмали для выявления аномалий в твердых тканях зубов. Анализируя сигналы Стокса и создавая изображения Мюллера, этот метод позволяет различать здоровые и больные ткани на основе их свойств ослабления и линейного замедления. Этот подход показывает перспективность неинвазивной диагностики стоматологических заболеваний, таких как кариес и синдром трещины зуба, в режиме реального времени [45].

Оптическое поляриметрическое обнаружение эффективно благодаря своей визуализации в реальном времени и неинвазивности, что делает его пригодными для повседневного клинического использования. Однако метод не может обеспечить такой же уровень молекулярной детализации, как, например, рамановская спектроскопия.

1.3.2.5 ИК-спектроскопия

В работе [7] проводилась диагностика пародонтита с использованием Фурье-ИК-микроскопии (FTIR) слюны и дискриминантного анализа методом частичных наименьших квадратов (PLS-DA). Исследовалось 22 образца слюны (10 пациентов с пародонтитом, 12 здоровых добровольцев). Точность классификации при перекрёстной проверке методом поэлементной кросс-валидации (LOOCV) составила 0,94.

Быстрый скрининг диабета и пародонтита с помощью FTIR-спектроскопии был предложен в статье [8]. Исследовались FTIR спектры слюны с применением методов машинного обучения. Метрикой качества являлась доля истинно положительных предсказаний (TPR): от 0,78 до 0,93 в зависимости от сравниваемых групп.

Достоинства:

- Чувствительность к органическим соединениям. Хорошо выявляет изменения в белках, липидах, углеводах – что важно при воспалениях, разрушении коллагена и метаболических сдвигах в тканях.
- Относительная простота и дешевизна оборудования. FTIR-спектрометры часто дешевле и проще в эксплуатации по сравнению с Рамановскими.
- Анализ сложных биологических образцов (например, слюны, десневой жидкости). Не требует длительной подготовки проб, может использоваться *in vitro* с минимумом пробоподготовки.
- Точность на уровне ROC AUC 0,9 и выше в комбинации с ИИ/ML. В исследованиях показывала точность до 0,94 при использовании машинного обучения.

Недостатки:

- Чувствительность к воде. Вода значительно поглощает ИК-излучение, что усложняет анализ живых тканей *in vivo* (например, десны и кости), где высокая гидратация.
- Меньшая пространственная разрешающая способность. Не позволяет детализированно анализировать клеточную и субклеточную структуру по сравнению с СКР.

1.3.3 Другие методы

1.3.3.1 Самодокладываемые (self-reported) диагностические опросники.

В статье [46] сообщается, что точность ROC AUC метода при выявлении пародонтита на 1-2 стадии составляет 0,61, а на 3-4 стадии 0,87. Причем в сегментах пациентов с определенной комбинацией возраста и курения значения метрик составили 0,95 ROC AUC, 0,96 чувствительности и 0,89 специфичности.

Преимуществами опросников является:

- Низкая стоимость и простота проведения. Опросники легко распространяются в больших выборках и используются для первичного скрининга.

- Удобство для проведения эпидемиологических исследований. Позволяют собрать большое количество информации от участников.

Недостатки:

- Низкая чувствительность. точность метода при выявлении ранних стадий пародонтита составляет порядка 0,6 ROC AUC;
- Субъективность данных. Результаты зависят от восприятия и памяти пациента, что снижает объективность.
- Ограниченная диагностическая точность. Высокая специфичность, но при этом значительное число ложноположительных или ложноотрицательных результатов.

1.3.3.2 Слюнные ферментные тесты

При использовании комбинации слюнных показателей (например, активности лактатдегидрогеназы, щелочной фосфатазы и выявления скрытой крови) одно из японских исследований продемонстрировало ROC AUC на уровне 0,76, чувствительность 0,9 и специфичность 0,62 (Kugahara и др., 2008) [47].

Преимущества:

- Неинвазивность.
- Быстрота анализа. Результаты можно получить в короткие сроки, что удобно для скрининга.
- Относительно невысокая стоимость. Не требует дорогостоящего оборудования, что позволяет проводить регулярный мониторинг.

Недостатки:

- Варьируемость специфичности. Имеется большое число ложноположительных результатов.
- Чувствительность к условиям сбора. Результаты могут зависеть от времени сбора слюны, состояния пациента (гидратация, время после еды) и метода обработки.

- Ограниченность маркеров. Не все биомаркеры в слюне могут быть использованы для точной диагностики, и их концентрации могут быть ниже пороговых значений для диагностики.

1.3.3.3 Энзиматический тест VANA (N-бензоил-DL-аргинина-2-нафтиламид)

Исследование, проведённое Туртоном и соавторами, показало, что данный тест обладает чувствительностью 0,86 и высокой специфичностью – 0,95, а также положительной прогностической ценностью 0,98 и отрицательной прогностической ценностью 0,73 при ROC AUC равном 0,90 (Turton и др., 2017) [48].

Преимущества:

- Высокая специфичность ~0,95.
- Хорошая чувствительность ~0,86, что позволяет выявить большинство случаев заболевания.
- Быстрота проведения. Тест может дать результат в течение минуты, что удобно для клинического скрининга.

Недостатки:

- Необходимость подготовки оператора. Для точного проведения теста требуется обучение и стандартизация процедуры.
- Ограниченность применения: Метод может быть специфичным для определённых групп пациентов (например, беременных), и его использование в широкой практике требует дополнительной валидации.
- Стоимость реактивов. Хотя сама процедура быстрая, стоимость реагентов может быть выше.

1.3.3.4 Методы на основе искусственного интеллекта (AI)

В статье [5] описывается разработка AI-моделей для анализа рентгенограмм и других изображений, при этом достигается общая точность диагноза примерно 0,94 при абсолютной чувствительности (1,0). Это позволяет сравнивать традиционные методы с современными подходами на базе искусственного интеллекта.

Преимущества:

- Высокая диагностическая точность около 0,94 и чувствительность 1,0 (при сравнении с экспертами).
- Снижение субъективности. Автоматизированное распознавание изображений минимизирует влияние человеческого фактора.
- Быстрая обработка. AI-системы могут анализировать большое количество изображений за короткий промежуток времени.
- Возможность интеграции. Такие системы могут интегрироваться в цифровые системы клиники, облегчая ведение документации и мониторинг лечения.

Недостатки:

- Высокая стоимость разработки и внедрения. Создание и поддержание AI-систем требует значительных инвестиций в программное обеспечение, аппаратные средства и обучение персонала.
- Зависимость от качества входных данных. Результаты напрямую зависят от качества рентгенограмм или других изображений.
- Необходимость контроля оператора. Несмотря на высокую чувствительность, существует риск ложноположительных результатов, что требует подтверждения диагнозов специалистом.

1.3.4 Спектроскопия комбинационного рассеяния

Ограничения перечисленных методов исследования твердых тканей зубов могут быть преодолены с помощью спектроскопии комбинационного рассеяния. Данный метод является простым, неинвазивным и быстрым способом для оценки тканей зубов [44] [41] [49] [50].

Авторы работы [41] с помощью спектроскопии комбинационного рассеяния исследовали процесс заживления и регенерации кости в процессе лечения пародонтита.

В статье авторов [44] описано использование рамановской спектроскопии для анализа структуры зубов по сравнению с синтетическими апатитами.

Обсуждаются спектральные линии, связанные с твердыми и мягкими тканями зуба, которые предоставляют важные данные для понимания химических структурных свойств дентина и эмали.

В контексте стоматологических исследований рамановская спектроскопия использовалась для изучения различных аспектов тканей зуба, включая эмаль и дентин. Этот метод особенно полезен для раннего выявления кариеса, оценки реминерализации и оценки воздействия отбеливающих средств и кислых продуктов на эмаль. Например, он может обнаруживать изменения в минеральном и органическом составе тканей зубов, предоставляя неинвазивный метод мониторинга здоровья зубов и эффективности стоматологического лечения [51].

В работе Mandra и др. [52] использовали рамановскую спектроскопию для оценки тканей зубов с кариесом. Исследования показывают, что это заболевание вызывает снижение концентрации ионов $(\text{PO}_4)^{3-}$.

В исследовании Gatin и др. (2024, Diagnostics) [41] изучали процессы заживления и регенерации костей при лечении пародонтита. Применена спектроскопия КР *in vivo* к альвеолярной кости и выявлено, что интенсивности фосфатных полос спектра коррелируют с качеством кости и статусом заболевания (пирофосфат указывает на патологические процессы). Авторы заключают, что спектроскопия КР может служить инструментом для оценки минерализации костной ткани при пародонтите. Хотя прямых численных метрик не приводится, работа демонстрирует жизнеспособность спектроскопии КР для анализа минерализованных тканей в клинике.

В работе [53], была продемонстрирована новая методика, основанная на поляризационной спектроскопии комбинационного рассеяния, для выявления ранних изменений эмали человека, вызванных искусственным кариесом. Авторы показали, что для здоровой эмали пик комбинационного рассеяния, возникающий из-за симметричного колебания PO_4^{3-} при 959 см^{-1} , был сильно поляризован. Это контрастирует со спектрами кариозных поражений, которые демонстрируют более слабую поляризационную зависимость при 959 см^{-1} . Эта разница в степени

анизотропии поляризации комбинационного рассеяния позволила отличить ранний кариес зубов от здоровой эмали.

В статье [54] с использованием традиционных рутинных гистологических методов и спектроскопии КР проведены сравнительные исследования минерализованных тканей челюсти человека и показано, что совместное использование этих методов позволяет получить значительно больше данных о патологических процессах в минерализованных тканях (в случае кариеса) для сбора, а также позволяющие определить особенности минерализации в условиях направленной костной регенерации.

В работе [9] использовалась спектроскопия КР в сочетании с алгоритмами машинного обучения (например, Extra Trees, AdaBoost, LDA) для различения трёх основных пародонтопатогенов: *Porphyromonas gingivalis* (Pg), *Fusobacterium nucleatum* (Fn) и *Aggregatibacter actinomycetemcomitans* (Aa). Заявлена точность классификации на уровне ROC AUC 0,94. Однако в работе не проводилась предобработка, выделение информативных признаков, не приведена интерпретация и полная оценка качества полученных классификаторов, которые могли быть сильно переобучены.

Экспериментальные исследования с использованием спектроскопии комбинационного рассеяния для оценки тканей зубов после кюретажа и процедур профессиональной гигиены полости рта в литературе не появлялись.

Спектроскопия комбинационного рассеяния (КР) обладает рядом преимуществ:

- Неинвазивность и быстрота: Метод позволяет проводить исследования без повреждения тканей и обеспечивает быстрый сбор данных, что особенно важно для динамического мониторинга.
- Высокая молекулярная детализация: в отличие от традиционных методов, КР предоставляет детальную информацию о химическом составе и структурных изменениях зубной ткани, что открывает возможность выявлять ранние патологические изменения.

- Портативность и адаптируемость: Технология может быть реализована в компактных диагностических устройствах, что делает её удобной для использования в клинических условиях.
- Отсутствие необходимости в подготовке образцов: В отличие от РЭМ и биохимического анализа, спектроскопия комбинационного рассеяния позволяет проводить исследования *in vivo*, исключая процедуру деструктивной подготовки.
- Работа с водной средой. Вода практически не мешает получению спектра – подходит для анализа живых тканей без дегидратации.
- Отсутствие вредного ионизирующего излучения. Спектроскопия КР использует лазерное освещение, которое не является ионизирующим, что делает метод безопасным для многократного применения и позволяет проводить мониторинг состояния тканей без риска воздействия на здоровье пациента.
- Возможность проведения анализа в реальном времени. Быстрая регистрация спектральных данных позволяет использовать метод для динамического наблюдения за изменениями в тканях во время процедур или в процессе лечения, что особенно ценно для оперативного принятия клинических решений.
- Высокая специфичность химического состава. Метод позволяет определять тонкие различия в химическом составе и структурных характеристиках тканей, что важно для раннего выявления патологий. Это предоставляет возможность проводить количественную оценку минерального и органического компонентов и устанавливать корреляции с патологическими изменениями.
- Универсальность применения. Спектроскопия комбинационного рассеяния может быть применена не только для диагностики стоматологических патологий, но и для анализа других биологических тканей. Такой междисциплинарный потенциал делает метод перспективным в более широком спектре биомедицинских исследований и клинической практики.

Недостатки:

- Наличие флюоресценции и случайных шумов может затруднять выделение истинных спектральных линий. Высокая чувствительность прибора к этим помехам требует применения специальных алгоритмов коррекции (например, методов декомпозиции спектров), что усложняет процесс анализа.
- Для корректного выделения информативных спектральных признаков необходимо проводить сложную предобработку (интерполяция, нормализация, сглаживание, коррекция базовой линии). Отсутствие стандартизированных алгоритмов может привести к переобучению моделей.
- Во многих исследованиях не проводилась комплексная оценка качества полученных классификаторов. Необходимость дополнительных исследований для интерпретации и валидации выделенных спектральных признаков указывает на то, что метод требует дальнейшей оптимизации для устойчивого клинического применения.
- Для получения качественных спектров необходимы современные рамановские спектрометры с лазерным источником, что может увеличить стоимость исследования. Кроме того, надежная калибровка и контроль параметров прибора являются важными условиями корректного измерения.
- Из-за большого числа этапов предварительной обработки и возможной вариативности спектральных данных требуется разработка и внедрение единых протоколов измерений и анализа.

1.4 Выводы по главе

В главе рассмотрены современные оптические методы анализа биоматериалов в стоматологии с акцентом на диагностику пародонтита и оценку состояния зубных тканей. Традиционные методы, такие как рентгенография, РЭМ, флюоресцентная и ИК-спектроскопия, несмотря на их высокую объективность и детализацию, имеют свои ограничения – необходимость использования ионизирующего излучения, деструктивность подготовки образцов, сложность

интерпретации полученных данных и невозможность выявления ранних патологических изменений.

Спектроскопия комбинационного рассеяния, напротив, демонстрирует значительный потенциал: она неинвазивна, обеспечивает высокую молекулярную детализацию и позволяет проводить динамический *in vivo* анализ без разрушительной подготовки образцов. Метод обладает рядом преимуществ, таких как возможность работы в водной среде, быстрота получения данных и безопасное применение благодаря отсутствию ионизирующего излучения.

Данная диссертационная работа направлена на проработку метода анализа спектров комбинационного рассеяния и разработку эффективных алгоритмов предобработки и декомпозиции спектров. Целью является решение выявленных выше проблем – уменьшение влияния флюоресцентного фона и шумов, стандартизация способа и повышение точности диагностической классификации. Результаты работы могут стать основой для создания инновационных диагностических систем, позволяющих проводить раннее выявление и динамический мониторинг стоматологических патологий.

2 ОБЗОР МЕТОДОВ ОБРАБОТКИ СПЕКТРАЛЬНЫХ ДАННЫХ КОМБИНАЦИОННОГО РАССЕЙЯНИЯ МНОГОКОМПОНЕНТНЫХ СРЕД

Цель данной главы – предоставить обзор существующих методов обработки спектральных данных комбинационного рассеяния (КР).

2.1 Спектроскопия комбинационного рассеяния и общая схема способа обработки данных

Спектроскопия КР применяется в качестве инструмента для мониторинга динамики процессов, происходящих в тканях, дает уникальные возможности изучения возбужденных состояний молекул, химических и биологических объектов.

Это простой в использовании, неинвазивный, неразрушающий метод. Он имеет высокое пространственное разрешение и информативность спектров, не требует специальной подготовки образцов, в процессе проведения экспериментов не оказывает механического, термического и радиоактивного воздействия на образец, что позволяет проводить многократные измерения.

Лазер с длиной волны возбуждения 785 нм вызывает комбинационное рассеяние в диапазоне 800–1200 (стоксовая компонента). Помимо комбинационного рассеяния от объекта имеется релеевское рассеяние на длине волны работы лазера 785 нм, которое на 6–7 порядков интенсивнее комбинационного рассеяния. Для подавления релеевского рассеяния, используется специально подобранный фильтр, подавляющий излучение на длине волны лазера на 5–6 порядков.

Полученный спектр отображает по оси ординат спектральную плотность интенсивности излучения из диагностического волокна в относительных единицах, а по оси абсцисс длину волны в нм. При этом изначальный спектр, имеет фиксированный шаг по оси абсцисс в нм, а преобразованный в обратные сантиметры [см^{-1}] имеет непостоянный шаг. Так, в области $\sim 380 \text{ см}^{-1}$ шаг составляет $1,92 \text{ см}^{-1}$, а при $\sim 1900 \text{ см}^{-1}$ шаг равен $1,48 \text{ см}^{-1}$, и в среднем $1,7 \text{ см}^{-1}$.

На рисунке 1 представлен спектр комбинационного рассеяния (КР) костной ткани. Типичный спектр КР биоматериала состоит из высокочастотных шумов, флуоресцентной составляющей и комбинационного рассеяния.

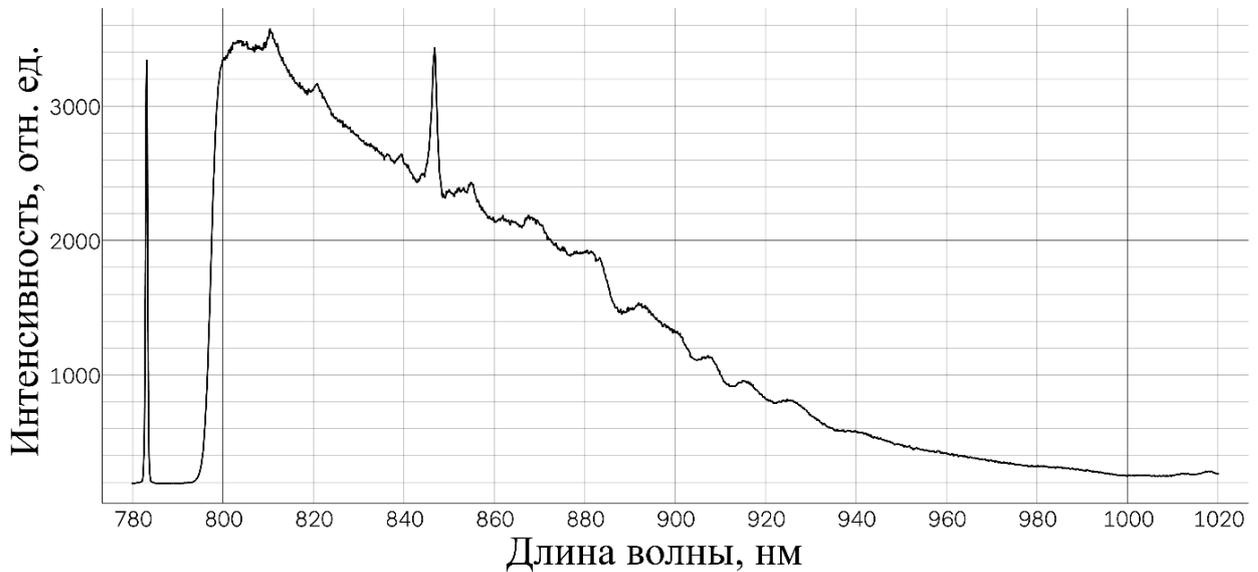


Рисунок 1 - Спектр комбинационного рассеяния костной ткани

2.1.1 Интерполяция

Применяется интерполяция спектров для приведения к одному диапазону и шагу длин волн. Наиболее адекватным является применение линейной интерполяции (рисунок 2), вычисляемое по формуле (1) для $x \in [x_0, x_1]$:

$$f(x) \approx y = P_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0), \quad (1)$$

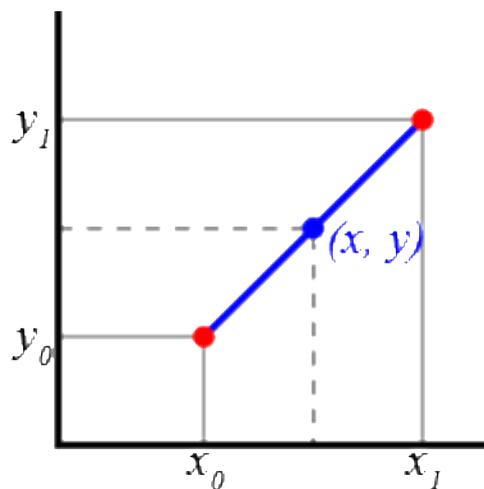


Рисунок 2 – Линейная интерполяция в точке x

2.1.2 Удаление случайных высокоинтенсивных шумов

Кроме высокочастотных случайных шумов в сигнале присутствуют высокочастотные шумы, наводимые с электросети (cosmic spikes), которые не ослабляются сетевым фильтром, с узкой шириной пика меньше ширины линии Релеевского рассеяния на длине волны возбуждающего лазера. Для подавления этих всплесков выбран алгоритм итеративного подавления всплесков, основанный на применении сглаживающего фильтра (Iterative mollifier-based despiking of the single-scan Raman spectra) [55].

2.1.3 Конвертирование единиц измерения длины волны

Рамановское смещение считается от длины волны упругого Релеевского рассеяния. В случае рисунка 1 используется возбуждающее излучение лазера на длине волны 785 нм. В спектроскопии КР для характеристики света вместо длины волны используются обратные сантиметры (измеряемое в см^{-1}). Что удобно в использовании, поскольку они линейно связаны с энергией.

Спектры комбинационного рассеяния регистрировались в диапазоне 780–1020 нм в виде двух столбцов с длиной волны и интенсивностью сигнала в относительных цифровых единицах. Переход к рамановскому смещению производится по формуле (2):

$$k(\lambda, \lambda_B) = \frac{10^7}{\lambda_L} - \frac{10^7}{\lambda + (\lambda_L - \lambda_B)}, \quad (2)$$

где: λ – значение длины волны, λ_L – длина волны излучения лазера, определенная при калибровке (истинная), λ_B – длина волны упругого рассеяния лазерного излучения фактическая (для текущего измерения). Данная формула позволяет компенсировать смещение длины волны при снятии серии спектров.

2.1.4 Нормализация

Следующим этапом обработки является нормализация спектров КР. В качестве алгоритма нормализации широко используется стандартизация (standard normal variate - SNV) [56], который реализуется с помощью формулы (3):

$$SNV_i = \frac{(I_i - \bar{I})}{\sqrt{\frac{\sum_i (I_i - \bar{I})^2}{n - 1}}}, i = 0, 1, 2 \dots n \quad (3)$$

где SNV_i – нормализованная интенсивность спектра в точке i , I_i – начальное значение интенсивности спектра в точке i , \bar{I} – средняя относительная интенсивность спектра, n – общее количество точек спектра.

Также для спектров КР может быть применен алгоритм нормализации расширенной мультипликативной коррекции сигнала (Extended multiplicative signal correction EMSC), который подробно описан в статье [57].

В дополнение к основным этапам предобработки, связанным с коррекцией и нормализацией базовой линии, EMSC также позволяет разделять и количественно определять различные типы химических и физических вариаций в колебательных спектрах. Таким образом, EMSC способствует достижению лучшей интерпретируемости колебательных спектров, а также делает последующие калибровочные модели более простыми и статистически более надежными.

Подход EMSC и его расширения можно адаптировать к стандартной схеме калибровки/валидации. В последние годы подход EMSC часто использовался для оценки и коррекции конкретных физических и химических характеристик как в спектрах инфракрасного поглощения, так и в спектрах комбинационного рассеяния, и ожидается, что в ближайшие годы этот подход разовьется в новых важных приложениях в колебательной спектроскопии.

Глобальная модель EMSC может быть рассчитана для спектров, включая наиболее важные главные компоненты, описывающие повторную коррекцию:

$$A(\tilde{\nu}) = a + \bar{x}(\tilde{\nu})b + d_1\tilde{\nu} + d_2\tilde{\nu}^2 + \sum_{k=1}^{NPC} g_k p_k(\tilde{\nu}) + e(\tilde{\nu}), \quad (4)$$

где, a – коэффициент базовой линии, \bar{x} – среднее значение спектра, b – коэффициент, описывающий оптический путь, принимаемый равным для всех компонентов образца, d_1 и d_2 – коэффициенты полинома базовой линии, g_k – значение матрицы экстинкции, NPC – количество главных компонент, p_k – коэффициент, относящийся к наиболее важным векторам повторной вариационной

нагрузки, оцененным с помощью метода главных компонент (МГК), e – остаточные различия спектров, имеющие смысл отличия концентрации компонентов образца.

Из (4) можно получить формулу для нормированного спектра КР:

$$A_{corr}(\tilde{\nu}) = \frac{A(\tilde{\nu}) - a - d_1\tilde{\nu} - d_2\tilde{\nu}^2 - \sum_{k=1}^{NPC} g_k p_k(\tilde{\nu})}{b}, \quad (5)$$

Также распространен алгоритм нормализации по площади под спектральной кривой с использованием трапецеидального правила, который заключается в том, чтобы сначала вычислить площадь под кривой (интеграл), а затем разделить каждое значение спектра на эту площадь. Это приводит к тому, что интеграл нормализованного спектра становится равен 1.

Для спектра $f(x)$, заданного в виде дискретных точек $(x_0, f_0), (x_1, f_1), \dots, (x_n, f_n)$, площадь под кривой A вычисляется как:

$$A = \sum_{i=0}^{n-1} \frac{(x_{i+1} - x_i)}{2} (f_i + f_{i+1}), \quad (6)$$

Это и есть правило трапеций (trapezoidal rule) – приближение интеграла как суммы площадей трапеций между каждой парой соседних точек. После расчета площади A , нормализованный спектр $f_{norm}(x_i)$ вычисляется по формуле:

$$f_{norm}(x_i) = \frac{f(x_i)}{A}, \quad (7)$$

Это интегральная нормализация: после неё вся площадь под спектром становится равной 1. Она применяется, чтобы учесть только относительное распределение интенсивностей, устранив влияние общего масштаба.

2.1.5 Сглаживание

Полученный сигнал спектра КР содержит шумы. Для сглаживания спектра КР могут быть использованы: фильтр Винера (Wiener filter) [58], Whittaker smoother [59], Savitzky-Golay filter [60], медианный фильтр, а также оконная фильтрация и другие.

Оконная фильтрация основана на свертке сигнала с оконной функцией. В качестве оконной функции используют треугольное, синусоидальное, окно

Ланцоша, Ханна, Хэмминга, Блэкмана, Харриса, Блэкмана-Харриса, прямоугольное окно, окно Наталла, Гаусса, Кайзера и множество других [61] [62].

Вариантом устранения шумов также является декомпозиция эмпирических мод (EMD - Empirical mode decomposition) [63]. С помощью EMD находятся высокочастотные составляющие спектра, которые в случае обработки спектров КР биоматериалов являются шумом и вычитаются из итогового спектра.

Алгоритм шумоподавления фильтр максимального правдоподобия Савицкого–Голея (Maximum Likelihood Estimation Savitzky-Golay filter MLE-SG) [64] построен на сочетании оценки максимального правдоподобия (MLE), основанной на оценке шума в спектре, со сглаживанием Савицкого–Голея (SG). Алгоритм пытается преодолеть классическую проблему обеспечения гладкого спектра без шума, не влияя при этом на точность резких интенсивных линий. Алгоритм состоит из двух конкурирующих ограничений, которые применяются к заданному спектру: первое условие, которое использует фильтрацию SG, предполагает, что спектр является гладким, т. е. данный образец не будет существенно отличаться от своих соседей; и второе условие, основанное на MLE, требует, чтобы выборка не отклонялась значительно от необработанного значения, которое было записано, принимая во внимание распределение шума, которое существует для этого необработанного значения выборки.

Данный алгоритм обеспечивает сглаживание спектра с сохранением резких спектральных линий, например линия на $\sim 960 \text{ см}^{-1}$, не требуется дополнительных настроек параметров пользователем. Также преимуществом алгоритма является использование отношения сигнал/шум для автоматической настройки степени сглаживания.

Далее могут быть применены алгоритмы коррекции базовой линии, декомпозиции спектра на элементарные линии. А полученные результаты могут быть проанализированы различными алгоритмами статистического анализа и машинного обучения.

Общая схема обработки спектров КР представлена на рисунке 3.

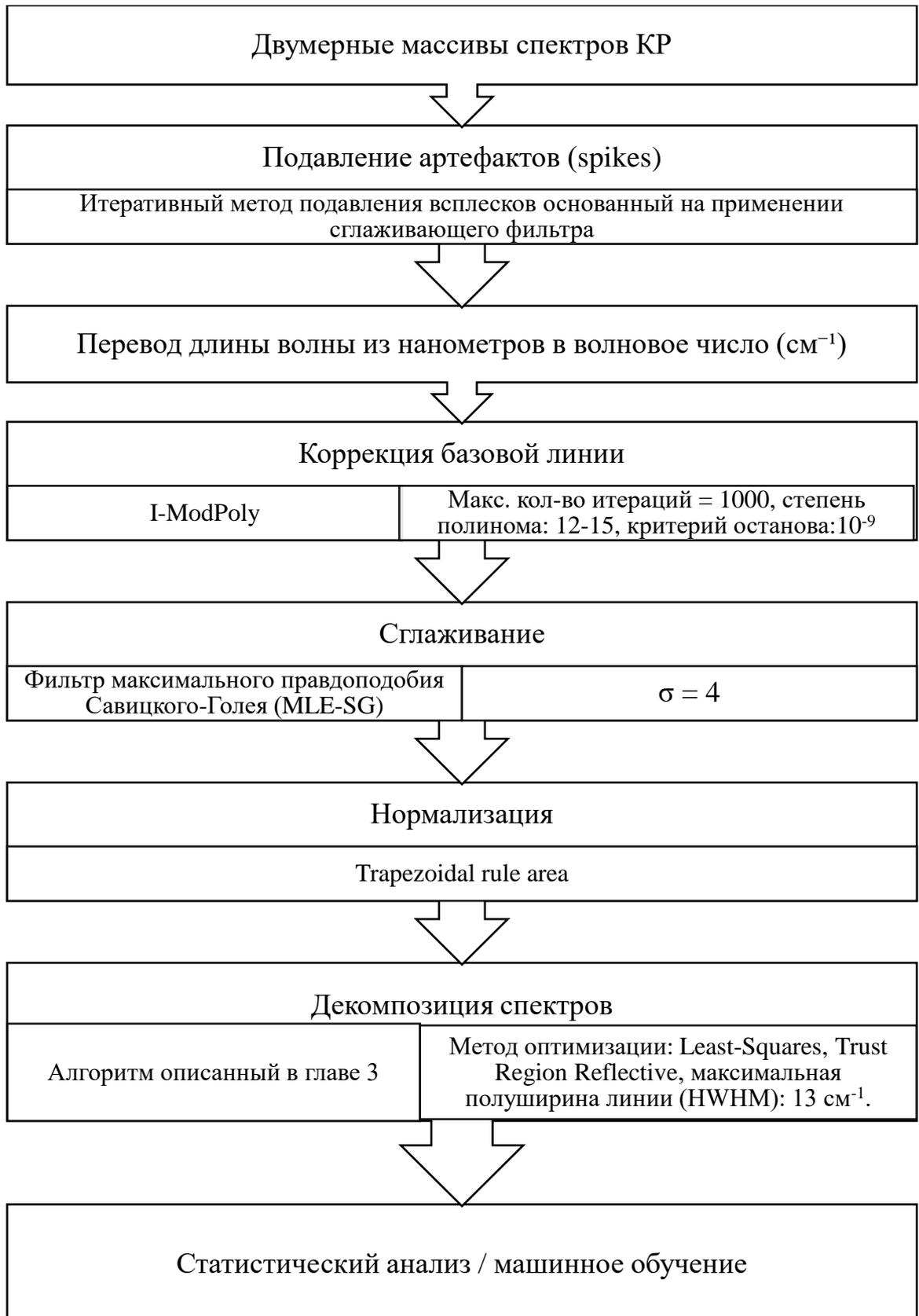


Рисунок 3 – Общая схема обработки спектров КР

2.2 Алгоритмы коррекции базовой линии спектров КР

Дрейф базовой линии (baseline drift) является часто встречающейся и серьезной проблемой в спектрах КР, особенно для биологических образцов. Основной причиной дрейфа базовой линии в спектроскопии КР является флюоресценция, генерируемая внутри образца. В ряде случаев интенсивность базовой линии может быть намного больше интенсивности линий КР, что может повлиять на последующий анализ спектра. Поэтому перед дальнейшим анализом необходимо удалить базовую линию. Если не проводить коррекцию базовой линии, то это повлияет на последующий качественный или количественный анализ [65] [66] [67] [68].

Сигнал в колебательной спектроскопии в основном состоит из химической информации, базовой линии и случайных шумов. Однако существование базовой линии и случайных шумов может отрицательно повлиять на качественные или количественные аналитические результаты, поскольку базовая линия всегда выглядит как независимая от выборки плавная кривая [69].

Наличие интенсивного флюоресцентного фона обусловлено наличием в анализируемой среде флуорофоров. Дальнейшая коррекция базовой линии является не просто вопросом эстетики; это влияет на качественную и количественную информацию, содержащуюся в наборе данных. В случае дальнейшего разложения спектра на элементарные линии такая коррекция становится обязательной.

В процессе обработки спектров КР используются различные алгоритмы коррекции базовой линии. Для исследования биоматериалов широко применяются алгоритмы итеративной аппроксимации полинома такие как Improved Modified Multi-Polynomial Fitting (улучшенный модифицированный алгоритм многополиномиального аппроксимирования I-ModPoly) [65].

В настоящее время для корректировки базовой линии в спектроскопии КР существует более 50 алгоритмов. Они делятся на:

Алгоритмы аппроксимации полинома:

- Modified polyfit (ModPoly) [66].
- Improved Modified Multi-Polynomial Fitting (I-ModPoly) [65]
- Locally estimated scatterplot smoothing (LOESS) [67].
- Quantile Regression [70].
- Goldindex Method [71].

Алгоритмы, основанные на Whittaker-smoother:

- Asymmetric Least Squares (AsLS) [72]
- Improved Asymmetric Least Squares (iAsLS) [73].
- Adaptive iteratively reweighted penalized least squares (airPLS) [69].
- Asymmetrically Reweighted Penalized Least Squares (arPLS) [74].
- Doubly Reweighted Penalized Least Squares (drPLS) [75].
- Improved Asymmetrically Reweighted Penalized Least Squares (iarPLS) [76].
- Adaptive Smoothness Penalized Least Squares (asPLS) [77].
- Peaked Signal's Asymmetric Least Squares Algorithm (PSALSA) [78].
- Derivative Peak-Screening Asymmetric Least Squares Algorithm (DerPSALSA) [79].

Морфологические алгоритмы:

- Morphological Penalized Least Squares (MPLS) [80].
- Morphological based (Mor) baseline algorithm [81].
- Improved Morphological based (IMor) baseline algorithm [82]
- Morphological and Mollified Baseline (MorMol) [83].
- Averaging Morphological and Mollified Baseline (AMorMol) [84].
- Rolling Ball Baseline [85].
- Morphology-Based Penalized Spline (MPSpline) [86].
- Joint Baseline Correction and Denoising (JBCD) [87].

Алгоритмы, основанные на аппроксимации сплайнов:

- Mixture Model [88].
- Iterative Reweighted Spline Quantile Regression (IRSQR) [89]
- Corner-Cutting Method [90]

Алгоритмы, основанные на сглаживании:

- Noise Median method [91].
- Statistics-sensitive Non-linear Iterative Peak-clipping (SNIP) [92].
- Small-Window Moving Average (SWiMA) [93].
- Iterative Polynomial Smoothing Algorithm (IPSA) [94].
- Range Independent Algorithm (RIA) [95].

Алгоритмы, основанные на классификации пиков и базовой линии

- Dietrich's Classification Method [96].
- Golotvin's Classification Method [97].
- Standard Deviation Distribution [98].
- FastChrom's Baseline Method [99].
- Continuous Wavelet Transform Baseline Recognition [68].
- Fully Automatic Baseline Correction [100].

И другие. Далее подробно разберем наиболее релевантные алгоритмы аппроксимации полинома в спектроскопии КР.

2.2.1 Алгоритм многополиномиального аппроксимирования ModPoly

Полиномиальная аппроксимация кривой имеет явное преимущество перед другими алгоритмами удаления фоновой флюоресценции в своей способности сохранять спектральные контуры. Однако простая аппроксимация полиномиальной кривой к необработанному спектру комбинационного рассеяния МНК не позволяет эффективно воспроизвести фон флюоресценции, поскольку аппроксимация будет основана на минимизации различий между базовой линией и измеренным спектром, который включает как фон флюоресценции, так и измеренный спектр, и пики комбинационного рассеяния. Таким образом, последующее вычитание этого подходящего полинома приводит к спектру, который изменяется относительно нулевой базовой линии. Поэтому крайне важно, чтобы полином соответствовал спектральным областям, содержащим только фоновую флюоресценцию, игнорируя спектральные области, содержащие полосы КР. Алгоритм полиномиальной аппроксимации традиционно основан на

выбранных пользователем спектральных линий, на которых основывается подбор. К сожалению, это субъективное вмешательство имеет несколько недостатков. Это отнимает много времени, так как пользователь должен обрабатывать каждый спектр отдельно, чтобы идентифицировать неактивные в спектре КР области спектра, которые будут использоваться при подборе полинома. Кроме того, идентификация неактивных в комбинационном рассеянии частот не всегда тривиальна, поскольку биологические спектры комбинационного рассеяния иногда содержат несколько соседних пиков или пиков, которые не сразу очевидны. Конечный эффект – алгоритм, подверженный изменчивости.

Также недостатком обычной полиномиальной аппроксимации является ее зависимость от диапазона спектральной аппроксимации и выбранного полиномиального порядка. В качестве примера на рисунке 4 [66] показан необработанный спектроскопический сигнал от кожи предплечья и подобранный полиномиальный фон флюоресценции при различных условиях.

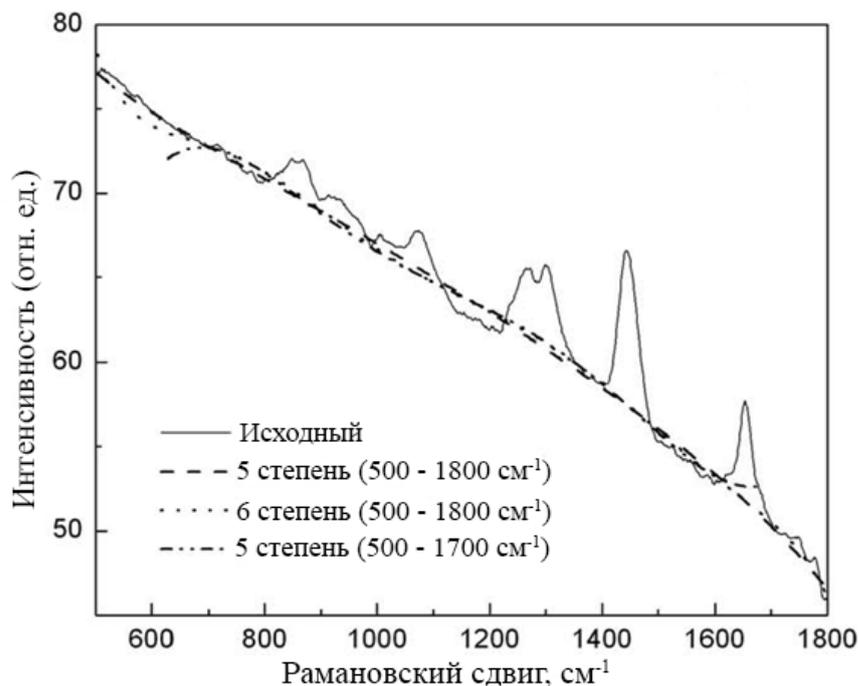


Рисунок 4 - Необработанный спектр КР кожи предплечья человека при различных условиях полиномиальной аппроксимации: одиночная полиномиальная аппроксимация с полиномами пятого и шестого порядка для диапазона данных от 500 см⁻¹ до 1800 см⁻¹ и полиномиальная аппроксимация пятого порядка для данных диапазон от 600 до 1700 см⁻¹

Результаты аппроксимации явно различаются в зависимости от порядка полинома для одного и того же спектрального диапазона (например, полином пятого и шестого порядка для диапазона 500–1800 см⁻¹), а также для разных диапазонов данных с использованием одного и той же степени полинома.

Чтобы устранить эти ограничения, был разработан модифицированный алгоритм подбора полинома для вычитания флуоресценции - modified polyfit method [66]. Этот алгоритм сглаживает спектр таким образом, что пики комбинационного рассеяния автоматически удаляются, оставляя нетронутой только базовую флуоресценцию, которую нужно вычесть из необработанного спектра.

В основе него лежит аппроксимация полинома кривой, основанная на МНК. Однако, чтобы исключить полосы КР из дальнейшей обработки, после первой итерации ограничиваются интенсивности линий КР, которые превосходят интенсивность подобранного полинома (рисунок 5).

Этот процесс (аппроксимация кривой и последующее перестроение) обычно повторяется от 25 до 200 итераций (в зависимости от таких факторов, как отношение флуоресценции к сигналу КР), постепенно удаляя высокочастотные пики КР из основной базовой флуоресцентной линии. В идеальном случае процесс фильтрации прекращается, когда на кривой соответствия больше нет точек данных, требующих переназначения (все значения равны или меньше соответствующих интенсивностей сглаженного спектра). Обработанный базовый спектр затем вычитается из необработанного спектра, чтобы получить полосы комбинационного рассеяния на почти нулевой базовой линии.

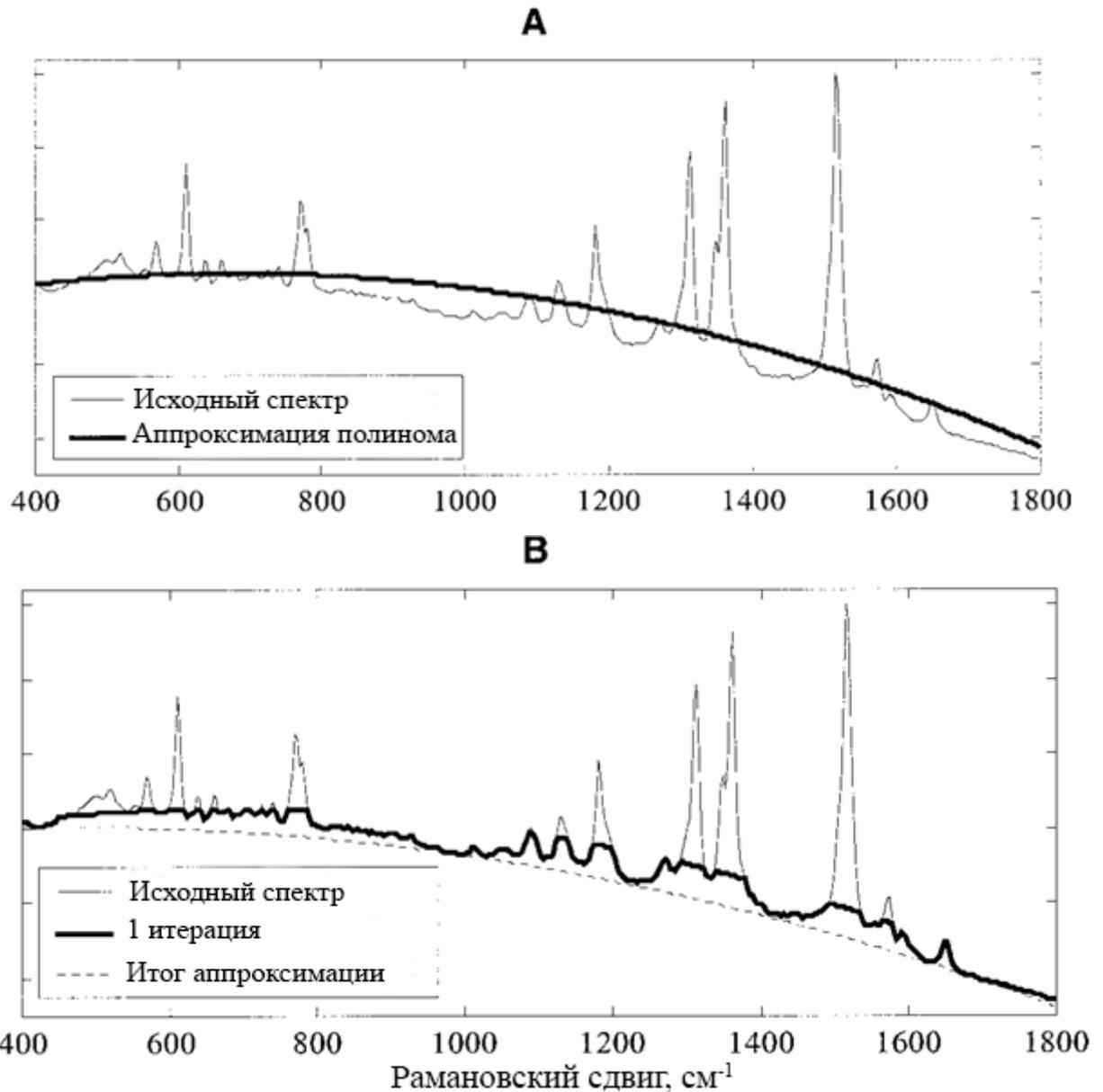


Рисунок 5 - Аппроксимация полиномиальной кривой применительно к родамину 6G. (А) Необработанный измеренный спектр КР и полиномиальная кривая, подобранная МНК, соответствует базовой флуоресценции. (В) Необработанный измеренный спектр КР, полиномиальная кривая после первой итерации и модифицированная сглаженная итоговая полиномиальная кривая вдоль основания измеренного спектра.

Алгоритм ModPoly существенно улучшает подавление флуоресценции в биомедицинских спектрах комбинационного рассеяния. Тем не менее, все еще существует ряд ограничений для автоматического удаления флуоресценции в реальном времени, которые необходимо учитывать:

- аналогично одиночному полиномиальному подбору, ModPoly может вводить искусственные пики, где исходные данные в областях без пиков немного выше кривизны полиномиальной кривой и, таким образом, неправильно заменяются подобранными;
- вклад некоторых больших пиков в полиномиальную аппроксимацию значителен, что может привести к смещению результатов аппроксимации.

2.2.2 Алгоритм многополиномиального аппроксимирования I-ModPoly

Для автоматического удаления флуоресценции субъективное прямое вмешательство человека должно быть сведено к минимуму, и, таким образом, проблемы с алгоритмом ModPoly должны быть решены для получения более репрезентативных спектров комбинационного рассеяния. Предложенный авторами статьи [65] алгоритм I-ModPoly (он же Vancouver Raman Algorithm) учитывает искажение сигнального шума и влияние больших пиков комбинационного рассеяния на аппроксимации фона флуоресценции. Он сочетает в себе процедуру удаления пиков во время первой итерации и учитывает эффект шума, который проявляется в алгоритме ModPoly.

Подробная схема алгоритма I-ModPoly показана на рисунке 6.

Он начинается с первой полиномиальной аппроксимации $P(\nu)$ с использованием необработанного сигнала КР $O(\nu)$, где ν - рамановский сдвиг. Затем остаток $R(\nu)$ и его стандартное отклонение DEV рассчитываются следующим образом:

$$R(\nu) = O(\nu) - P(\nu) \quad (8)$$

$$DEV = \sqrt{\frac{(R(\nu_1) - \bar{R})^2 + (R(\nu_2) - \bar{R})^2 + \dots + (R(\nu_n) - \bar{R})^2}{n}} \quad (9)$$

где n – общее количество отсчетов спектра КР.

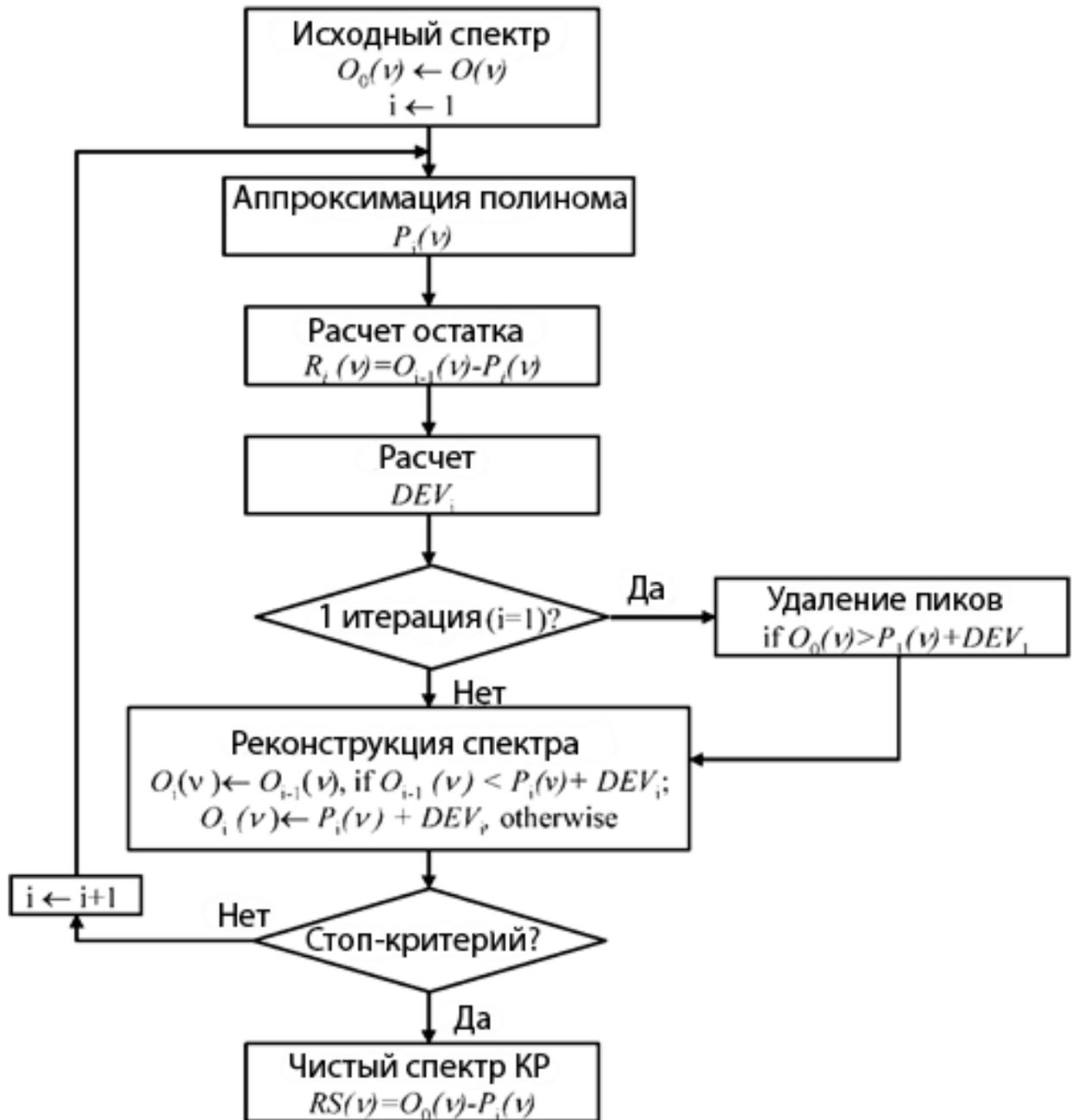


Рисунок 6 – Блок-схема алгоритма I-ModPoly

2.2.3 Другие алгоритмы коррекции базовой линии

Вариационная модовая декомпозиция (VMD) [101] разлагает спектр на различные собственные функции моды (IMF), что позволяет изолировать высокочастотные компоненты шума. Основным преимуществом VMD является его способность решать задачи смешивания мод и эффекты конечных точек лучше, чем традиционные алгоритмы, такие как эмпирическое модовое разложение (EMD) и дискретное вейвлет-преобразование (DWT) [102]. Этот алгоритм особенно

эффективен для сохранения острых пиков в спектрах, которые часто теряются при использовании других алгоритмов шумоподавления.

Еще один алгоритм Neural Network-Based Iterative Curve-Fitting использует подход, основанный на нейронной сети, для итеративной аппроксимации кривой и коррекции базовой линии [103]. Этот подход динамически выбирает основу функции в соответствии с трендом базовой линии, обеспечивая более точную подгонку, особенно для спектров с нерегулярной формой базовой линии. Модель нейронной сети повышает точность за счет обучения на данных моделирования для более эффективной обработки шума, чем традиционная полиномиальная аппроксимация и адаптивные МНК с итеративным перевзвешиванием (PLS).

2.3 Обзор алгоритмов декомпозиции спектров КР

Для спектров КР биоматериалов характерна высокая степень перекрытия линий со сплошным спектральным контуром. Существующие алгоритмы для декомпозиции спектров КР направлены на спектры неорганической природы без существенного перекрытия линий. Для разрешения перекрывающихся полос предложено два условно разделяемых принципиальных подхода – алгоритмы деконволюции и косвенные алгоритмы жесткого моделирования.

Были предложены слепые и полуслепые алгоритмы деконволюции. Как правило, слепые и полуслепые алгоритмы деконволюции используют априорные знания об инструментальной функции уширения для построения параметрической модели [104], а некоторые из них используют различные регуляризаторы [105] [106].

Слепые и полуслепые алгоритмы деконволюции позволяют в определенной степени восстановить исходные спектры и оценить инструментальную функцию уширения. Однако нельзя полностью разрешить перекрывающиеся полосы и избежать влияния шума, поэтому гладкость восстановленных спектров по-прежнему оставляет желать лучшего.

Алгоритм непрямого жесткого моделирования Альсмейера [107] [108] устанавливает пикообразную модель спектров комбинационного рассеяния для

разрешения перекрывающихся полос и часто используется, но использование модели профиля суммарной линии Гаусса-Лоренца (Pseudo-Voigt line), представляющей спектры комбинационного рассеяния в этом алгоритме, не поддерживается никакой физической теорией, на что уже указывал Мейер [109] и по сути является аппроксимацией функции Фойгта (Voigt) с ошибкой в 1 % [110]. В этой статье указывается что действительный спектр КР представляет собой набор линий Лоренца, а измеренный спектрометром спектр КР – набор функции Фойгта (Voigt). Также указывается что правильный подбор числа линий, их ширины и позиции крайне важен и является ключевым фактором успешного моделирования спектрального контура.

Voigt функция – это результат конволюции распределения Коши-Лоренца и распределения Гаусса. Алгоритмы деконволюции спектра как раз направлены на подбор функции уширения, которая в общем случае описывается распределением Гаусса. Эффекты уширения любого происхождения приводят к свертке лоренцева профиля линии и профиля, характерного для уширения [109].

Уширение с помощью профиля линии Гаусса подразумевает, что необходимо умножить лоренцев профиль на гауссовский профиль, что приводит к так называемому профилю Фойгта, который представляет собой свертку (произведение) гауссова и лоренцева профилей:

$$V(\nu) \propto \int_{-\infty}^{\infty} L_0 \frac{\omega_L^2}{\nu - \nu' + i\omega_L} G_0 \exp\left[-\left(\frac{\nu'}{\omega_G}\right)^2\right] d\nu', \quad (10)$$

где L_0 – амплитуда линии Лоренца, G_0 – амплитуда линии Гаусса, ν – длина волны, ω – ширина линии.

Алгоритм, описанный в статье [111] моделирует спектр КР с помощью Voigt функций, которая позволяет также учитывать уширение линии Лоренца. Но других принципиальных различий, кроме выбора функции линии, от алгоритмов [107] [108] и других аналогичных нет.

Большинство существующих алгоритмов декомпозиции спектров применимы для спектров неорганических веществ без сложного перекрытия множества линий, в большинстве случаев, определяющие состав линий только по

одному измеренному спектру. Причем созданная модель имеет жестко фиксированные параметры положения и ширины линий.

Алгоритм разложения, реализованный в статье [112], эффективно объединяет эволюционные алгоритмы с оптимизацией на основе градиента для решения сложной проблемы спектрального разложения. Инновационное использование GGA повышает производительность за счет введения разнообразия в генетические операции, тем самым повышая надежность решения.

Традиционные алгоритмы, такие как сглаживание Савицкого-Голея (SG) и дискретное вейвлет-преобразование (DWT), широко использовались для спектрального шумоподавления и разложения. Однако эти алгоритмы иногда не позволяют точно восстановить истинный сигнал, особенно при наличии шума или сложных спектральных характеристик [113] [114].

Недавние достижения представили более сложные подходы, такие как разложение вибрации Гильберта (HVD) и алгоритмы на основе глубокого обучения. HVD, например, разлагает спектры КР на компоненты, что позволяет более точно идентифицировать пики и восстанавливать их из зашумленных данных. Этот алгоритм показал значительные улучшения по сравнению с традиционными с точки зрения сохранения сигнала и снижения шума [113].

Расширенные алгоритмы спектрального разложения, особенно те, которые интегрируют фреймворки машинного обучения и глубокого обучения, предлагают несколько преимуществ. К ним относятся улучшенная точность идентификации пиков, улучшенное шумоподавление и способность обрабатывать большие наборы данных со сложными спектральными характеристиками. Например, применение сверточных нейронных сетей (CNN) показало себя многообещающим в уточнении начальных оценок, предоставляемых традиционными алгоритмами, что привело к более точным спектральным реконструкциям. Согласно исследованию, представленному в [113], [114], [115], модель CNN была обучена на разнообразном наборе данных спектров КР. Этот подход продемонстрировал замечательную точность в разложении сложных спектральных данных с перекрывающимися пиками и различными уровнями шума. Способность CNN обобщать данные

обучения позволила ей эффективно обрабатывать новые, неизвестные спектры с минимальным ручным вмешательством.

Однако эти расширенные алгоритмы также имеют ограничения. Требование обширных наборов данных обучения и вычислительных ресурсов может быть существенным препятствием, особенно для исследователей с ограниченным доступом к таким ресурсам. Разработка алгоритмов разложения не ограничивается только спектроскопией КР. Такие методы, как спектроскопия ядерного магнитного резонанса (ЯМР), масс-спектрометрия и хроматография, также в значительной степени зависят от эффективного спектрального разложения для точного анализа. Например, в ЯМР-спектроскопии для разложения и интерпретации сложных спектров использовались такие методы, как максимальная энтропия и байесовский подход. Аналогично, в масс-спектрометрии такие алгоритмы, как деконволюция перекрывающихся пиков с использованием функций Гаусса и Лоренца, являются общепринятой практикой [116].

Эти методы, хотя и эффективны в своих областях, подчеркивают универсальность, необходимую в подходах спектрального разложения. Каждый метод представляет уникальные проблемы и преимущества, часто требуя индивидуальных решений для достижения оптимальных результатов.

Априорная информация о количестве линий, их положениях и ширине очень важна, но не всегда доступна. Количество линий, используемых в модели, всегда должно быть минимальным, если только нет конкретных доказательств того, что следует использовать больше линий. Причина этого заключается в том, что, когда из общей формы огибающей не ясно, что требуется больше пиков, оптимизация модели с большим количеством линий вносит дополнительную неоднозначность в результат [109].

Например, работе [117] для определения примерного количества, положений, амплитуды и ширины линий используется метод производных с использованием производных 1-4-го порядка.

Функция Гаусса моделируется как:

$$z_j(\lambda) = A_j \exp\left(-\frac{(\lambda - \bar{\lambda})^2}{2\sigma_j^2}\right), \quad j = \overline{1, N}, \quad (11)$$

где A_j – амплитуда линии, $\bar{\lambda}$ – координата максимума, σ_j^2 – дисперсия.

Производные порядков $l = 1, 2, 3, 4$ от $z_j(\lambda)$ по λ равны:

$$z_j'(\lambda) = -\frac{A_j}{\sigma_j} \exp\left(-\frac{(\lambda - \bar{\lambda})^2}{2\sigma_j^2}\right) \left(\frac{\lambda - \bar{\lambda}}{\sigma_j}\right), \quad (12)$$

$$z_j''(\lambda) = -\frac{A_j}{\sigma_j^2} \exp\left(-\frac{(\lambda - \bar{\lambda})^2}{2\sigma_j^2}\right) \left(-\frac{(\lambda - \bar{\lambda})^2}{\sigma_j^2} + 1\right), \quad (13)$$

$$z_j'''(\lambda) = \frac{A_j}{\sigma_j^3} \exp\left(-\frac{(\lambda - \bar{\lambda})^2}{2\sigma_j^2}\right) \left(-\frac{(\lambda - \bar{\lambda})^2}{\sigma_j^2} + 3\right) \left(\frac{\lambda - \bar{\lambda}}{\sigma_j}\right), \quad (14)$$

$$z_j^{IV}(\lambda) = \frac{A_j}{\sigma_j^4} \exp\left(-\frac{(\lambda - \bar{\lambda})^2}{2\sigma_j^2}\right) \left(\frac{(\lambda - \bar{\lambda})^4}{\sigma_j^4} - 6\frac{(\lambda - \bar{\lambda})^2}{\sigma_j^2} + 3\right), \quad (15)$$

Из формул (11)–(15) следует, что

$$\begin{aligned} z_j'(\bar{\lambda}_j) &= 0, & z_j''(\bar{\lambda}_j) &= -\frac{A_j}{\sigma_j^2} < 0 \\ z_j'''(\bar{\lambda}_j) &= 0, & z_j^{IV}(\bar{\lambda}_j) &= 3\frac{A_j}{\sigma_j^4} > 0 \end{aligned} \quad (16)$$

Отрицательный экстремум 2-й производной и положительный экстремум 4-й производной соответствуют максимуму j -й линии, а нулевое значение 1-й и 3-й производных (при условии 16) также указывают на j -ю линию. Амплитуду и ширину линий можно примерно оценить как:

$$A_j = 3 \frac{z_j''(\bar{\lambda}_j)^2}{z_j^{IV}(\bar{\lambda}_j)}, \quad \sigma_j = \sqrt{3 \frac{-z_j''(\bar{\lambda}_j)}{z_j^{IV}(\bar{\lambda}_j)}}, \quad j = \overline{1, N}, \quad (17)$$

Ближайшим аналогом алгоритму, описанному в 3 главе, можно считать функцию "Guess Peaks" в программе MagicPlotPro, предназначенную для автоматического обнаружения пиков в спектре. Она работает следующим образом:

- Программа вычитает базовую линию из данных и применяет сглаживание с использованием алгоритма Савицкого–Голея.

- После предварительной обработки программа ищет глобальные экстремумы в данных и добавляет пики в этих точках. Каждый новый пик добавляется последовательно после вычитания предыдущих.

Этот алгоритм эффективен для быстрого и приблизительного определения положений пиков, но он имеет ограничения при работе с перекрывающимися пиками или при наличии значительного шума в данных.

2.4 Выводы по главе.

Рассмотрена общая схема обработки спектров комбинационного рассеяния, состоящая из следующих этапов: удаление высокочастотных шумов, конвертация единиц измерения, коррекция базовой линии, нормализация, сглаживание, и декомпозиция спектра на элементарные линии.

Проанализирована применимость существующих алгоритмов коррекции базовой линии и декомпозиции спектров КР.

Большинство существующих алгоритмов декомпозиции спектров КР позиционируются как алгоритмы для спектров неорганических веществ без сложного перекрытия множества линий, определяющие состав линий только по одному измеренному спектру. Причем созданная модель имеет жестко фиксированные параметры положения и ширины линий.

Данный обзор алгоритмов обработки спектров КР сформировал теоретическую основу для разработки модифицированного алгоритма декомпозиции, о котором пойдет речь в следующей главе.

3 РАЗРАБОТКА АЛГОРИТМА ДЕКОМПОЗИЦИИ СПЕКТРОВ КОМБИНАЦИОННОГО РАССЕЯНИЯ БИОМАТЕРИАЛОВ

Цель данной главы – разработка нового алгоритма декомпозиции спектров КР биоматериалов на совокупность элементарных компонент (линий КР), описываемых гауссовыми/лоренцевыми/Voigt-профилями.

Для достижения поставленной цели решаются следующие задачи:

1. Построение формальной модели КР-спектра в виде суммы элементарных профилей (Гауссовых, Лоренцевых и Voigt), способной воспроизводить ключевые особенности экспериментальных спектров при отсутствии артефактов.

2. Разработка и реализация итеративного алгоритма оптимизации состава компонент и их параметров (центральных частот, амплитуд и ширин).

3. Исследование влияния шумовых и фоновый составляющих на стабильность оценок параметров: применение алгоритмов количественной оценки неопределённости (например, доверительных интервалов) для параметров элементарных профилей.

4. Формулирование критериев оптимальности декомпозиции, включающих:

- минимизацию среднеквадратичной ошибки аппроксимации (MSE) и максимизацию R^2 ;
- парсимонию модели, то есть предпочтение более простой модели при сопоставимой способности описания данных, оцениваемую с помощью информационных критериев (AIC, BIC);
- соответствие параметров элементарных компонент физическим ожиданиям и литературным данным.

5. Валидация алгоритма на экспериментальных данных спектров КР дентинных и костных образцов: анализ чувствительности алгоритма к изменениям уровня шума.

Решение перечисленных задач позволит создать воспроизводимый и автоматизированный инструмент для анализа сложных КР-спектров, применимый в биомедицинских исследованиях и клинической практике.

3.1 Общая схема алгоритма

Входными данными является массив анализируемых спектров КР после процедуры коррекции базовой линии и параметры, устанавливаемые пользователем.

Для устранения вклада флуоресценции в спектр КР могут применяться различные алгоритмы коррекции базовой линии.

Единственным требованием к алгоритму коррекции базовой линии является определение базовой линии, близкой к истинной, чтобы она не опускала контур спектра в отрицательную область или не поднимала его. Если спектр КР определить как кривую 0 на рисунке 7, то в процессе разложения спектра могут появиться линии, описывающие не существующие в реальности элементарные линии в диапазоне $\sim 300 \text{ см}^{-1}$.

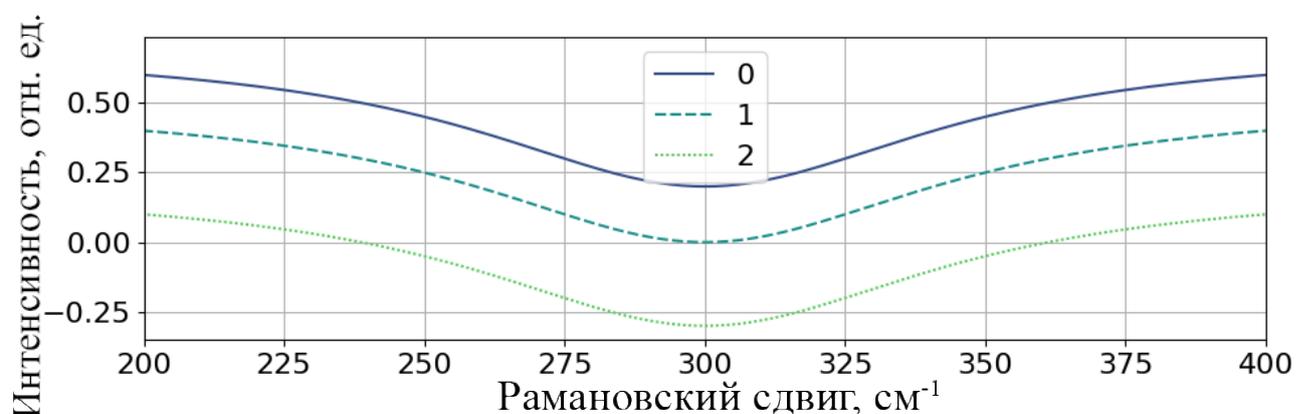


Рисунок 7 – Результат определения базового скорректированного спектра комбинационного рассеяния в области локального минимума

А в случае кривой 2 часть спектра КР находится в отрицательной области. Предпочтительный случай показан как кривая 1.

На рисунке 8 показана общая блок-схема нового разработанного алгоритма декомпозиции спектров КР, состоящего из двух основных алгоритмов:

итерационного подбора состава линий для каждого спектра КР и поиска общей модели, удовлетворяющей всем проанализированным спектрам.



Рисунок 8 – Блок-схема предлагаемого алгоритма декомпозиции спектра

Результатом работы алгоритма является модель, состоящая из набора элементарных линий с диапазонами значений параметров описывающих форму этих линий. Далее подробно описаны этапы работы алгоритма.

3.2 Описание алгоритма итеративного поиска состава линий

Алгоритм основан на последовательном добавлении количества линий и аппроксимации полученной модели пока добавление новых линий существенно не улучшит модель. Разберем основные принципы работы алгоритма на примере спектральной области спектра комбинационного рассеяния костной ткани $400 - 516 \text{ см}^{-1}$ (рисунок 9).

Данный спектральный контур имеет несколько линий КР с высоким коэффициентом взаимного перекрытия, так что определить количество, положения и ширину этих линий затруднительно.

Предположим, что расстояние между двумя соседними максимумами элементарных линий не меньше наибольшей половины полуширины линии HWHM .

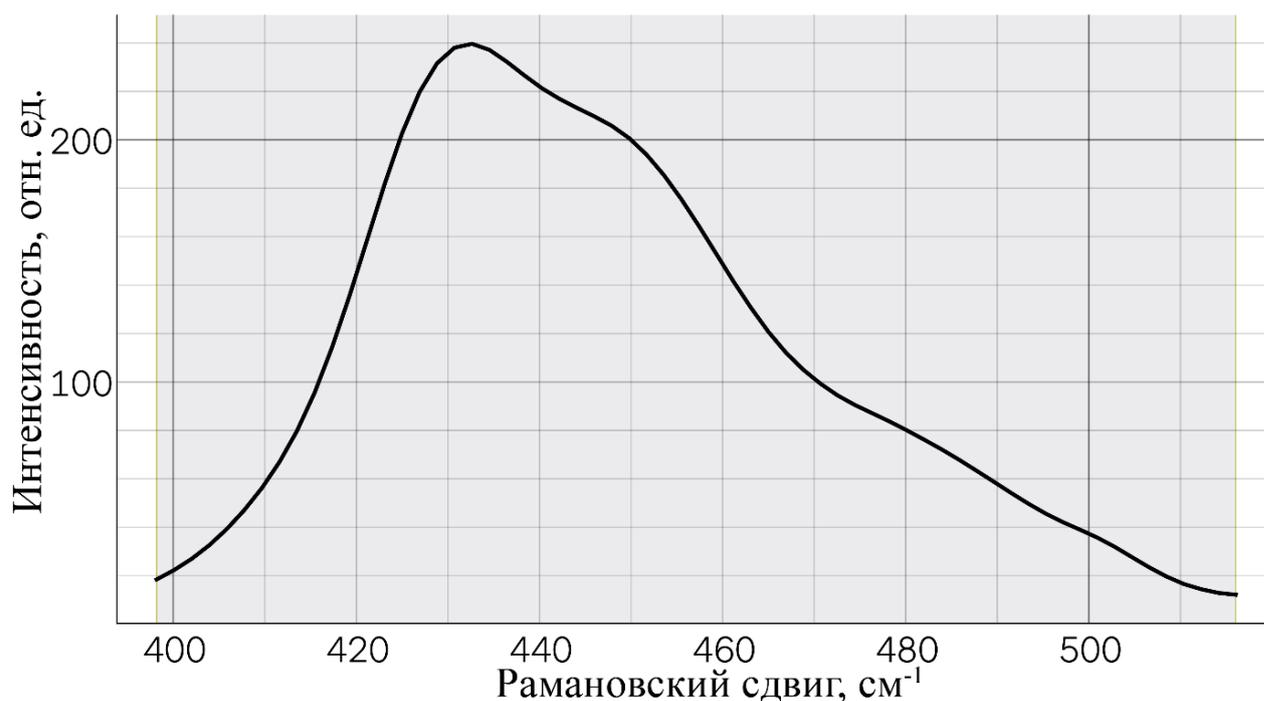


Рисунок 9 – Область спектра КР в диапазоне $400 - 516 \text{ см}^{-1}$

Вначале найдем глобальный максимум спектра. На первой итерации максимумом является точка с координатами $432,3 \text{ см}^{-1}$. Именно там в диапазоне $\pm 1 \text{ см}^{-1}$ будет находиться максимум первой линии. В качестве элементарной линии выберем Гауссиан. Оптимизатор – МНК. В качестве элементарной линии также может быть использован Лоренциан, *PreudoVoigt* или *Voigt*.

На рисунке 10 показан результат добавления одной линии без ограничения ширины линии. Видно, что подобранная линия выходит за рамки исходного спектрального контура и в дальнейшем результат добавления последующих линий будет значительно отличаться от физически обоснованного состава и распределения линий.

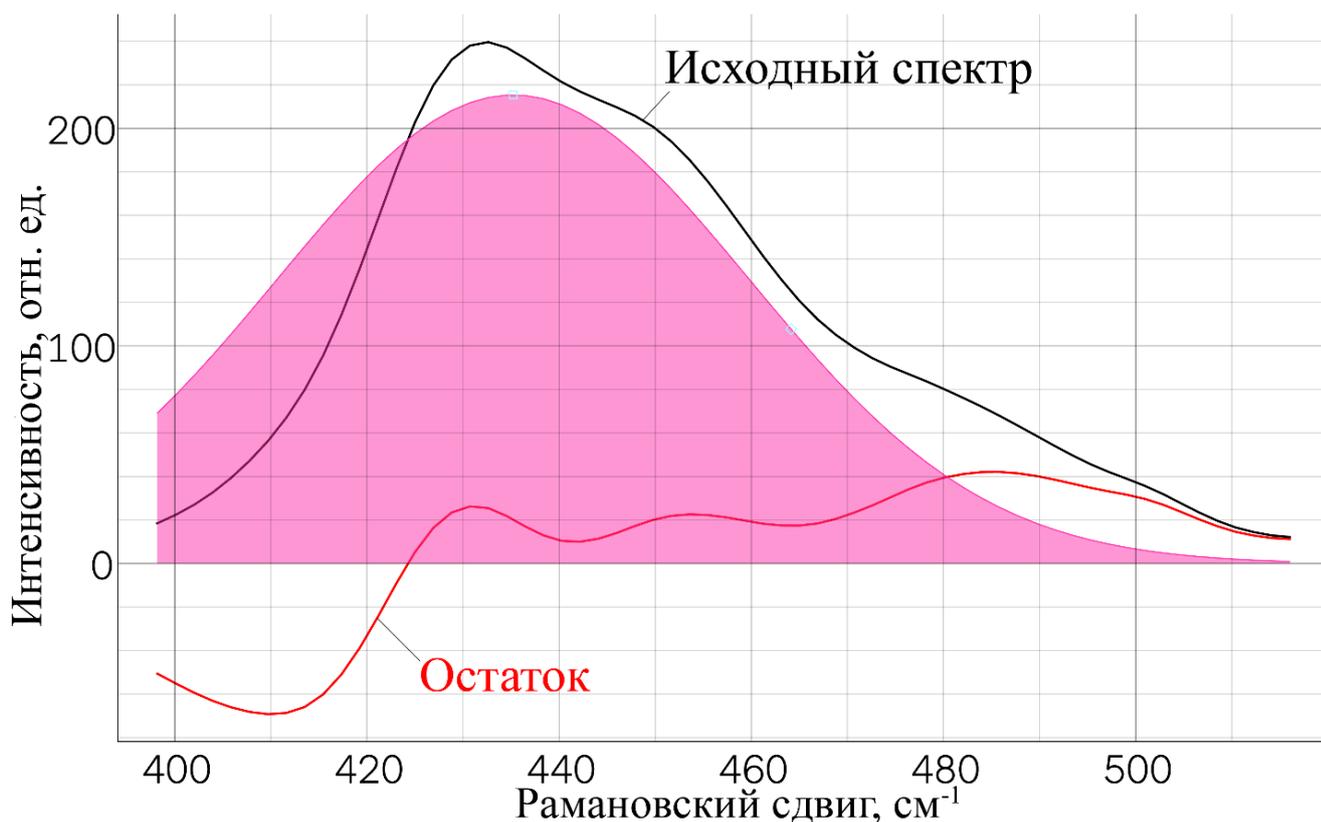


Рисунок 10 – Добавление первой линии без ограничения ширины линии

Поэтому в качестве входного параметра указывается максимально возможная полуширина линии на уровне половины амплитуды (HWHM) всех линий. На рисунке 11 показан результат с ограничением полуширины линии в 12 см^{-1} . Видно, что теперь подобранная линия не выходит за границы спектрального контура.

При выполнении декомпозиции ширину пиков не фиксированная, что приводит к более точному определению параметров линий.

На второй и последующих итерациях положение следующей линии определяется по максимуму остатка, здесь максимум второй линии находится в точке $\sim 456 \text{ см}^{-1}$. Результат второй итерации алгоритма показан на рисунке 12. Видно, что вторая добавленная линия выходит за пределы спектрального контура,

что будет минимизировано в последующих итерациях. Если в процессе добавления новая линия окажется на месте уже добавленной на предыдущих итерациях, то эта новая линия пропускается с обнулением области в $\pm \frac{1}{2}$ НWHM от положения линии.

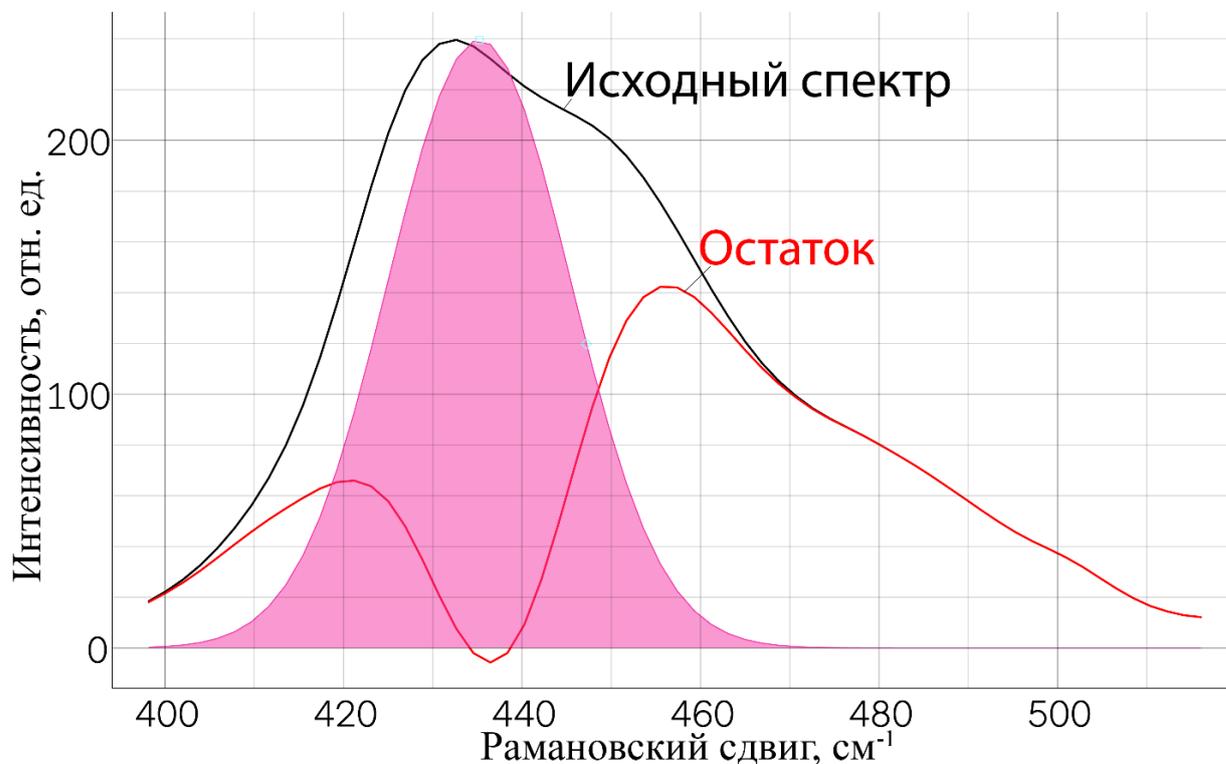


Рисунок 11 – Добавление первой линии с ограничением НWHM = 12 см⁻¹

Итоговый результат подбора модели, описывающей спектральный контур показан на рисунке 13. Лимитирующей амплитудой новой линии выбрано значение в 12,82 отн. ед. для ограничения бесконечного добавления новых линий. Амплитуда ниже 12,82 (значение изменяется пользователем) считается шумом, который аппроксимировать не нужно.

Так как для разных спектров подобранная модель для одного и того же контура может отличаться, нельзя считать одну из полученных моделей итоговой и требуется анализ для нахождения модели, подходящей под все спектры КР. Если анализируемый спектр один, то подобранная модель является итоговой.



Рисунок 12 – Результат добавления второй линии

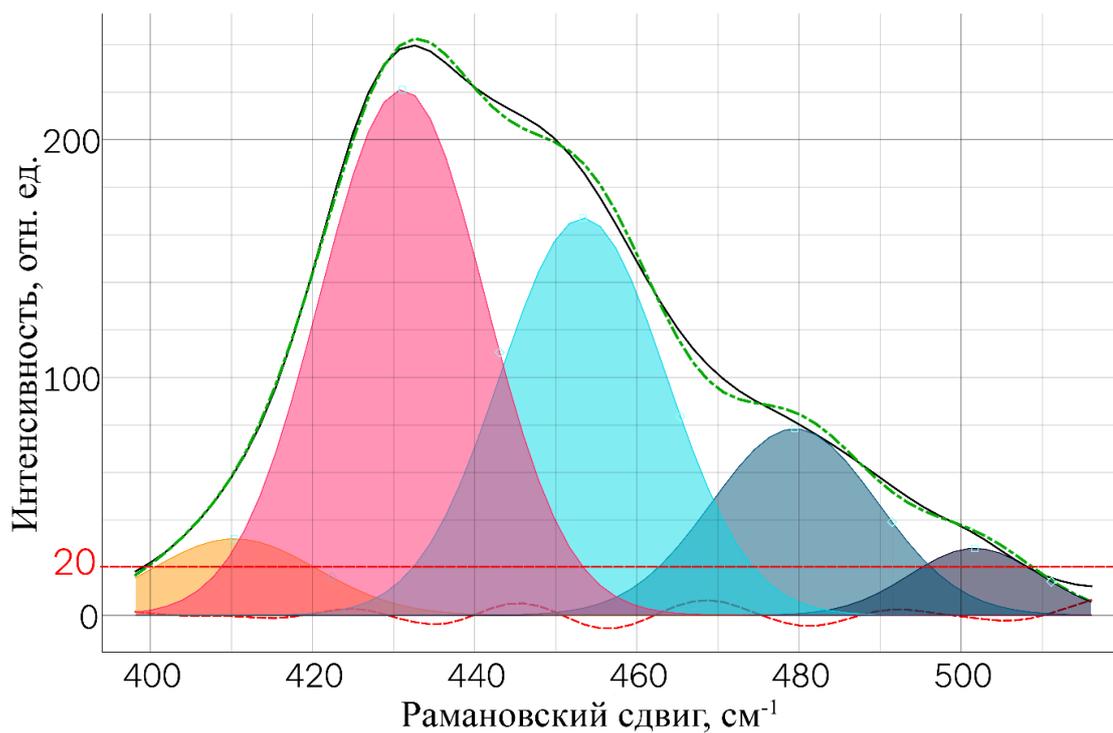


Рисунок 13 – Итоговый результат добавления линий для модели спектрального контура

Блок схема алгоритма итеративного поиска модели описывающей спектральные контуры КР показана на рисунке 14.

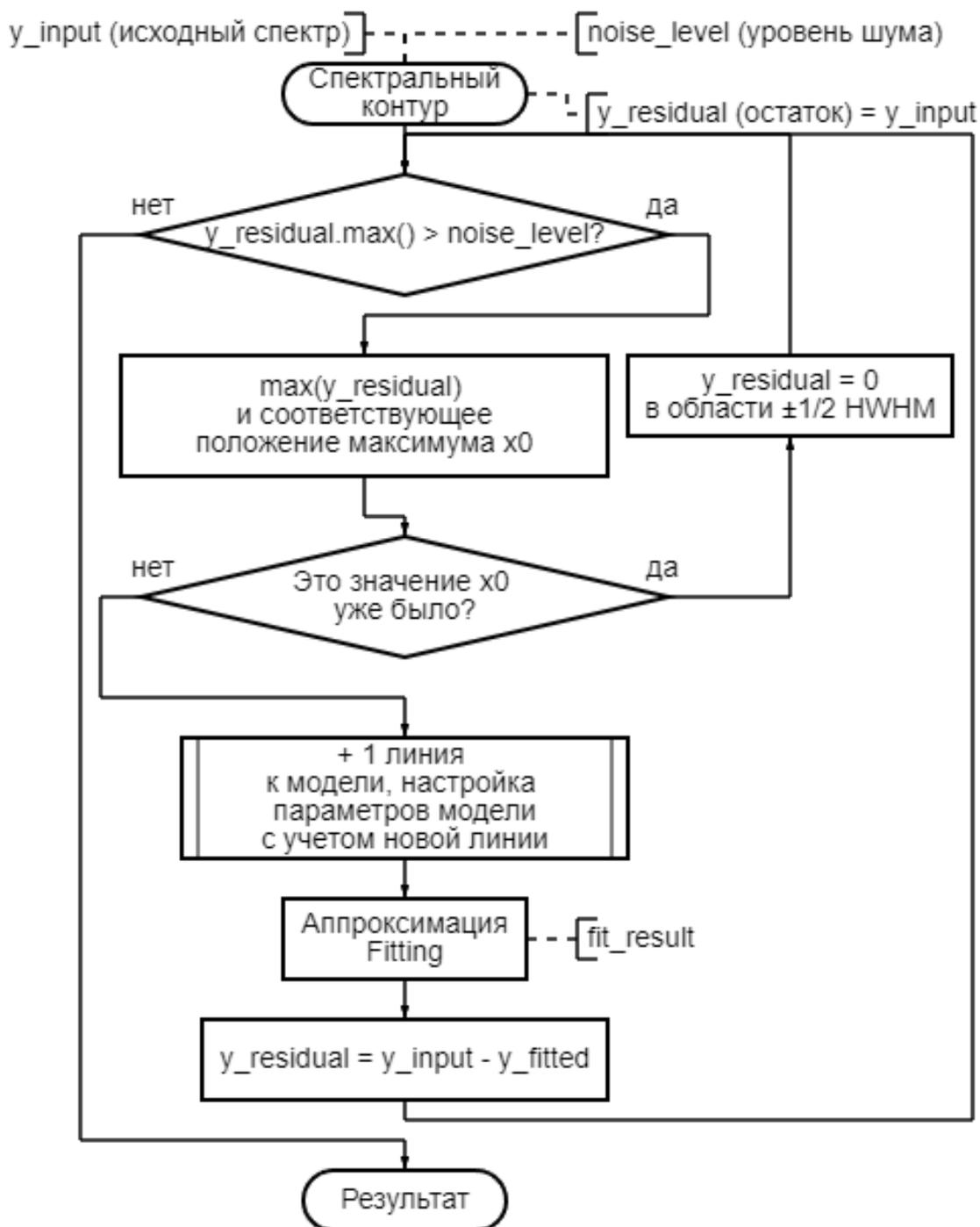


Рисунок 14 – Блок-схема итерационного алгоритма подбора модели, описывающей спектральный контур

В качестве входных параметров используется спектральный контур (x_input , y_input) и параметры, устанавливаемые пользователем:

- 1) Метод оптимизации модели. Предлагаемый алгоритм декомпозиции не привязан к конкретному методу оптимизации. Можно выбрать один из:
 - МНК, метод отражения доверительной области (Least-Squares, Trust Region Reflective method) [118];

- Алгоритм Левенберга – Марквардта (Levenberg-Marquardt) [119] [120];
- Метод Нелдера – Мида (Nelder-Mead) [121];
- Алгоритм Бройдена – Флетчера – Гольдфарба – Шанно ограниченный по памяти (Limited-memory BFGS - L-BFGS-B) [122];
- Метод Пауэлла (Powell) [123];
- Метод сопряжённых градиентов (Conjugate-Gradient) [124];
- Ограниченная оптимизация с помощью линейной аппроксимации (Constrained optimization by linear approximation Cobyla);
- Алгоритм Бройдена – Флетчера – Гольдфарба – Шанно (BFGS) [125];
- Усеченный метод Ньютона (Truncated Newton) [126];

И этот список может быть расширен.

- 2) Максимальная полуширина линии НВНМ max_dx .
- 3) Пороговое значение амплитуды линии $noise_level$, при достижении которой прекращается добавление новых линий.
- 4) Опционально для линии Voigt и Pseudo-Voigt. Ограничение максимальной доли распределения Лоренца по отношению к распределению Гаусса.

Далее использован метод наименьших квадратов (МНК) «Least-Squares, Trust Region Reflective», $max_dx = 12.0$, $noise_level = 0.0074493$.

Для аппроксимации спектра в данном случае выбрана реализация МНК с использованием библиотеки *lmfit*, поскольку он позволяет гибко задавать границы и начальные значения параметров и устойчив к флуктуациям фона. Алгоритм обеспечивает высокую точность подгонки ($R^2 > 0,99$ в большинстве случаев), а также устойчив к шуму при разумном выборе стартовых параметров.

В цикле происходит добавление новых линий к модели, оптимизация модели, обновление остатка $y_residual$. В случае если при начале работы алгоритма уже

добавлены линии пользователем, то анализ происходит с учетом этих начальных параметров.

Заранее количество линий и параметры линий неизвестны.

На параметры линии накладывается ряд ограничений:

- Амплитуда a не больше максимального значения в интервале $x_0 \pm \text{NWHM}$ и не меньше 0.
- NWHM dx , ограничен сверху выбором пользователя max_dx , снизу полушириной линии на 0 см^{-1} (на длине волны возбуждающего лазера).
- Колебания значений положения линии x_0 рассчитываются по формуле:

$$\Delta x_0 = \frac{\text{min_fwhm}}{4} + 1, \quad (18)$$

Эти ограничения не позволяют линии выйти за границы исходного спектрального контура, обеспечивая тем самым устойчивость, сходимость и воспроизводимость решения в сравнении с распространённым «свободным» моделированием контура. Также если происходит повторение определения координаты добавляемой линии, то область, прилежащая к этой координате, обнуляется для предотвращения бесконечного цикла.

Результатом анализа алгоритмом является модель-сумма линий, описывающая каждую спектральную область. Соответственно, в этой модели есть информация о положении каждой линии и количестве этих линий в конкретной области спектра. Но ни количество линий, ни их параметры еще не могут считаться окончательными и требуется проанализировать модели полученные для все обрабатываемых спектров КР, что происходит на следующем втором этапе алгоритма.

Также алгоритм может учитывать заранее добавленные пользователем линии КР на основе априорной информации. Тогда новые линии КР будут добавляться к уже существующим.

3.3 Анализ моделей после итерационного поиска состава линий

Результатом анализа на первом этапе (алгоритм, описанный в пункте 3.2) является набор координат положений линий по каждой спектральной области каждого анализируемого спектра. Варианты моделей декомпозиции преобразуются в массивы координат x_0 положений линий. Например, для первой спектральной области $380 - 516 \text{ см}^{-1}$ набор x_0 для N спектров может выглядеть так:

$$\begin{aligned} x_1 &= [430.38, 452.91, 475.71, 409.45, 388.46, 498.37] \\ x_2 &= [430.21, 452.91, 475.68, 410.36, 386.74, 504.03, 491.37] \\ x_3 &= [431.07, 452.98, 477.86, 388.65, 405.45, 498.57, 417.92, 515.68] \\ x_N &= [\dots\dots] \end{aligned}$$

На этом этапе еще точно неизвестно итоговое количество линий, координаты центра кластера и границы колебаний координаты. При дальнейшем описании для примера ограничимся данными первых трех спектров КР.

Для формирования итоговой модели из множества моделей используется алгоритм, блок-схема которого представлена на рисунке 15.

Первой операцией является кластеризация с использованием случайно выбранных двух спектров в качестве первых кластеров.

Блок-схема алгоритма кластеризации показана на рисунке 16. На вход алгоритма подается массив координат x_0 положений линий $x0_lines = [x_1, x_2, x_3]$, максимальное пороговое значение полуширины линии $HWHM$. Функция кластеризации вызывается N раз (количество спектров КР).

Так как начальные значения кластеров еще не известны, то сформируем первые кластеры на основе двух первых массивов x_1 и x_2 с помощью алгоритма кластеризации DBSCAN [127] с параметрами $\epsilon = HWHM = 12$, $min_samples = 2$.

В результате кластеризации DBSCAN будут получены кластеры:

$$\begin{aligned} c_1 &= [430.21, 430.38], c_2 = [452.91, 452.91], c_3 = [475.68, 475.71], \\ c_4 &= [410.36, 409.45], c_5 = [386.74, 388.46], c_6 = [504.03, 491.37, 498.37] \end{aligned}$$

Если всего анализировалось 2 спектра, то на этом кластеризация закончена и возвращается арифметическое среднее каждого кластера.

Если анализируется более двух спектров, то все остальные наборы x_i обрабатываются следующим образом.

Рассчитываются центры всех кластеров:

$$x_means = [430.29, 452.91, 475.695, 409.905, 387.6, 497.92]$$

Для первой точки 431.07 из x_3 расстояния до центров кластера будет:

$$distances = [0.78, 21.84, 44.625, 21.165, 43.47, 66.85]$$

Следовательно, так как расстояние до кластера c_1 наименьшее и меньше $NWHM = 12$, то точка 431.07 добавляется в кластер c_1 и среднее значение кластера c_1 обновляется. Эта процедура повторяется для всех остальных точек. Для точки 515.68 расстояния до всех центров превышают 12, и точка 515.68 образует новый кластер c_7 [515.68]. Цикл завершается после добавления точек всех наборов x_i . Результат после добавления точек из x_3 :

$$\begin{aligned} c_1 &= [430.38, 430.21, 431.07], c_2 = [452.91, 452.91, 452.98], \\ c_3 &= [475.71, 475.68, 477.86], c_4 = [409.45, 410.36, 405.45, 417.92], \\ c_5 &= [388.46, 386.74, 388.65], c_6 = [498.37, 504.03, 491.37, 498.57], \\ c_7 &= [515.68] \end{aligned}$$

Функция кластеризации с формированием начальных кластеров из $[x_1, x_2]$, вернет такой же результат, как и $[x_1, x_3]$ с усредненными значениями кластеров:

$$z_1 = z_2 = [430.55, 452.93, 476.42, 410.80, 387.95, 498.09, 515.68]$$

Но если для определения начальных кластеров использовать x_2 и x_3 , то результат незначительно изменится, так как точка 515.68 окажется в одном кластере с 504.03 и количество кластеров уже будет другим:

$$z_3 = [430.55, 452.93, 476.42, 387.95, 410.80, 501.604]$$

В качестве наборов для формирования начальных кластеров должно использоваться как можно больше вариантов. Так как при данном алгоритме кластеризации необходимо учитывать, что если первые кластеры сформированы на основе двух девиантных наборах координат x_0 , то количество кластеров в результате может изменяться, как и положение определенных линий.

Поэтому процедура кластеризации повторяется N раз и каждый раз первые 2 набора данных для определения начальных кластеров, к которым далее будут добавляться точки, выбираются случайным образом. Тогда в формировании

первых кластеров участвует $1 - 1/e^2 \approx 86,46\%$ от всей выборки. Полученный набор средних значений кластеров является множеством размером N и содержит возможные варианты итоговой модели. Последующие этапы показаны в рисунке 14, после первого цикла.

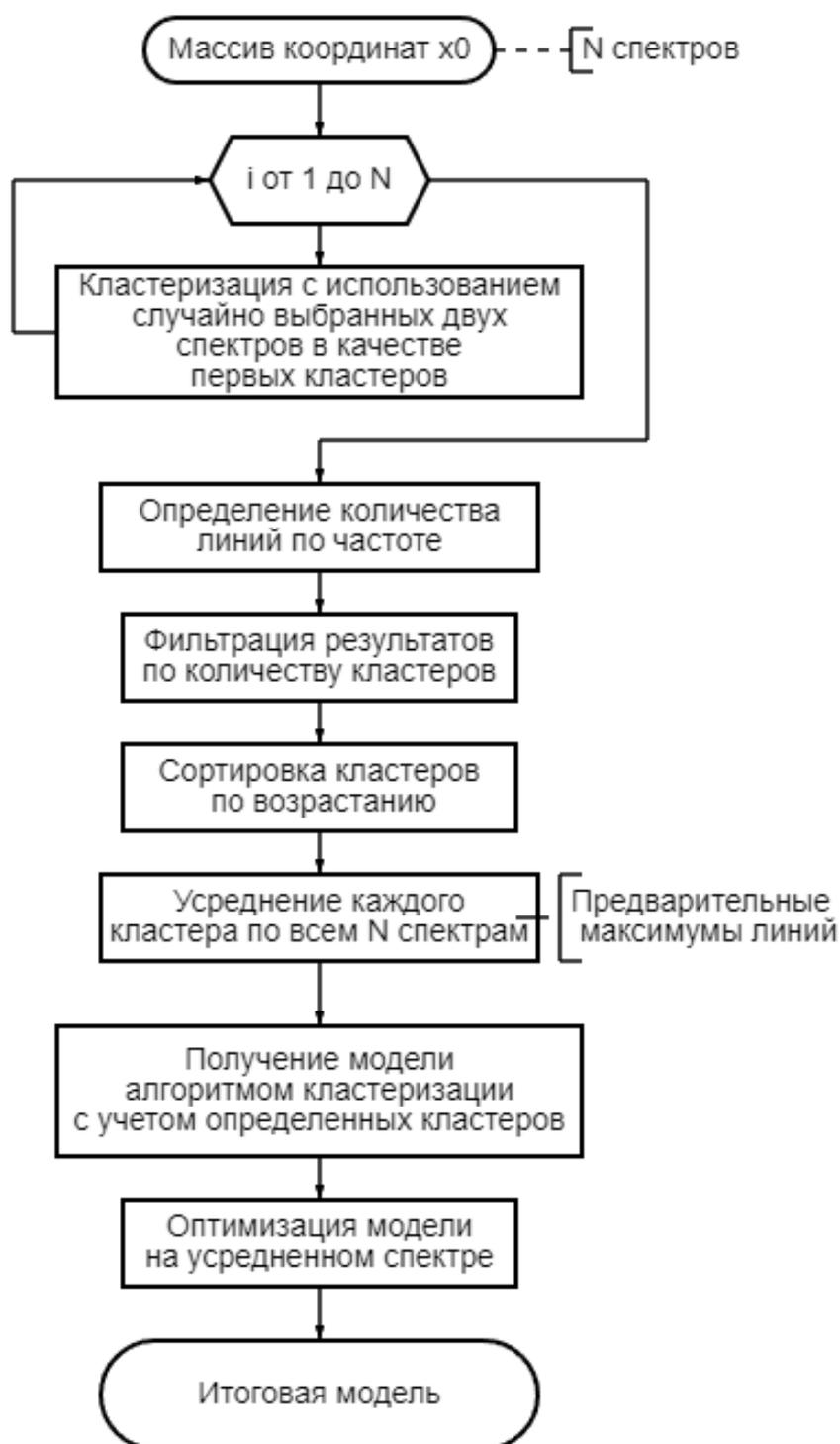


Рисунок 15 – Блок-схема алгоритма нахождения модели декомпозиции спектров КР

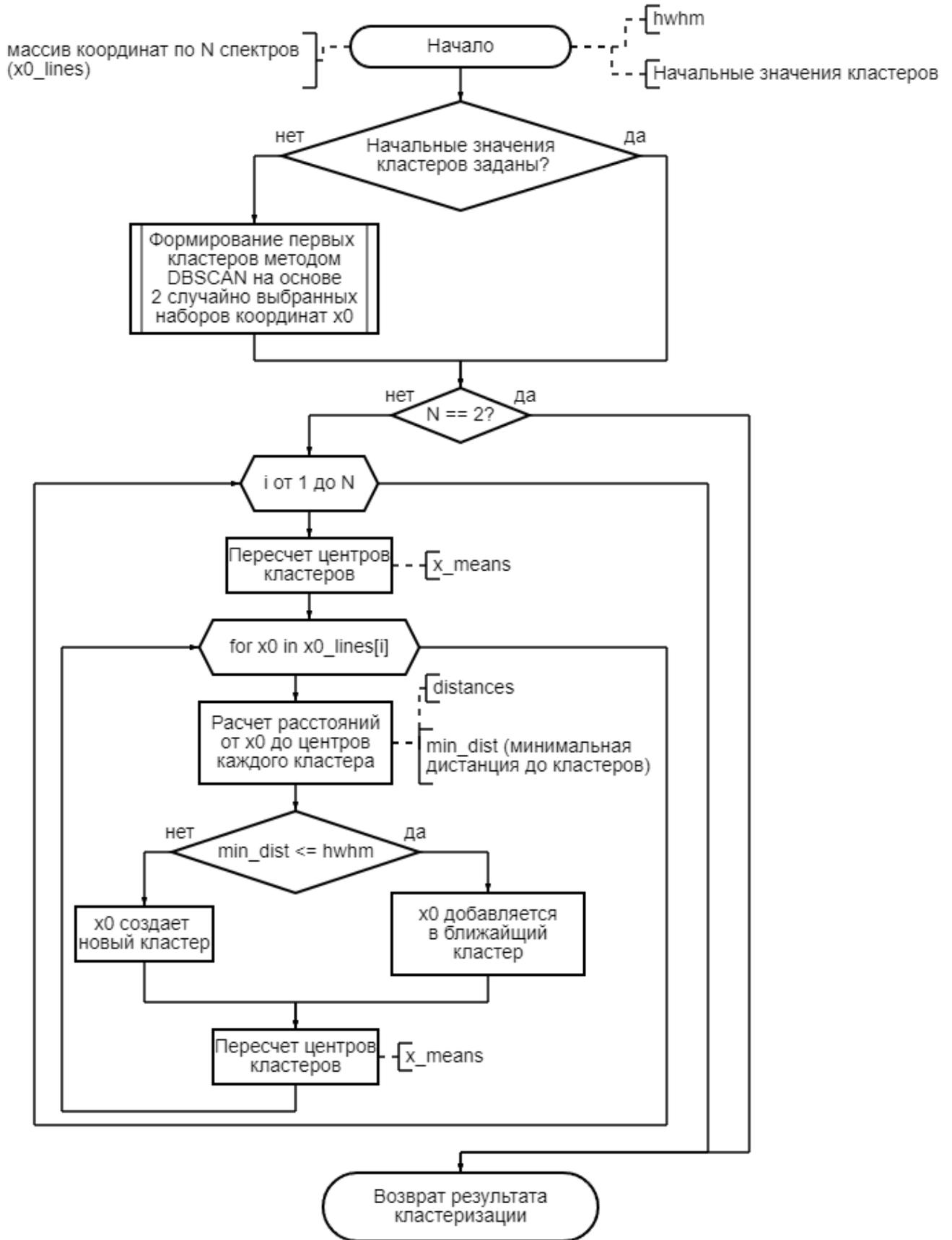


Рисунок 16 – Блок-схема алгоритма кластеризации

Далее происходит **определение количества линий по частоте**. Наиболее часто встречаемое количество линий (мода) в данном примере равно 7, так как в z_1 и z_2 количество элементов, а в z_3 6.

Все результаты **фильтруются** по этому количеству линий, то есть x_3 удаляется из дальнейшего анализа. Оставшиеся массивы z_1 и z_2 сортируются по возрастанию: $z_1 = z_2 = [387.95, 410.80, 430.55, 452.93, 476.42, 498.09, 515.68]$.

Далее происходит усреднение всех значения каждого кластера для получения начальных положений линий. Например, при получении отсортированных массивов одинакового размера:

[388, 411, 431, 453, 476, 498, 516]

[389, 413, 429, 452, 474, 499, 515]

[386, 410, 430, 455, 477, 499, 518]

Результатом усреднения будет один массив:

$z = [387.6, 411.3, 430, 453.3, 475.6, 498.6, 516.3]$

Полученный набор опять подается в алгоритм кластеризации (рисунок 16) в качестве начальных кластеров, без их формирования на основе случайно выбранных с помощью DBSCAN. Тогда для x_1, x_2, x_3 при найденном z результат кластеризации с вычисленными стандартными отклонениями будет:

$[387.86 \pm 0.76, 410.90 \pm 4.04, 430.42 \pm 0.4, 453.03 \pm 0.16, 476.21 \pm 0.95, 498.19 \pm 4.02, 516.0 \pm 0.31]$

Полученные средние кластеров являются центром области координат линии КР, а стандартное отклонение всех точек кластера является отклонением центра x_0 . На основе этих данных (количество линий, области положения координат линий) производится итоговая оптимизация модели на усредненном спектре КР. И полученный результат является готовой моделью для всех анализируемых спектров КР.

Разработанный алгоритм, по сути, анализирует плотность распределений возможных вариантов положений линий в спектре КР и возвращает составленную модель для декомпозиции.

На рисунке 17 показана плотность распределения найденных координат линий для каждой области спектра на основе анализа 344 спектров КР дентина и найденные положения x_0 со стандартным отклонением Δx_0 .

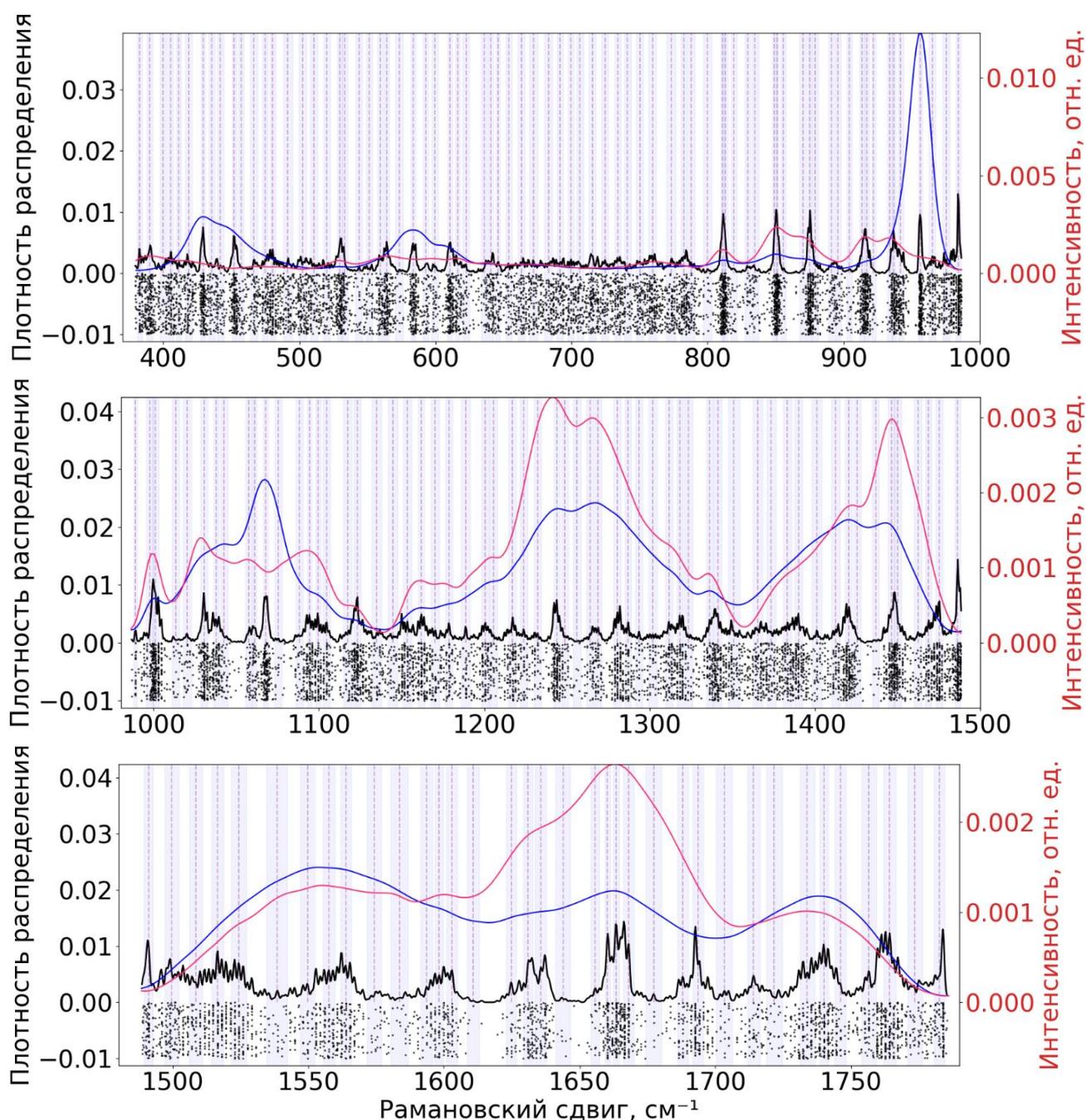


Рисунок 17 – Плотность распределения (черная линия) найденных координат линий для каждой спектральной области 344 спектров КР дентина.

Синяя линия – усредненный спектр минерализованного дентина, красная – деминерализованного, синяя полупрозрачная полоса – границы найденных элементарных линий

Видно, что определенные положения элементарных линий совпадают с максимумами плотности распределения и находятся в местах нахождения локальных максимумов спектрального контура.

На рисунке 18 показана модель декомпозиции, состоящая из 87 гауссиан, оптимизированная для усредненного спектра КР.

Чтобы интерпретировать спектр на рисунке 18, мы должны учитывать наличие областей, которые могут происходить из аморфного материала, которые часто бывают очень широкими. Аморфные материалы обычно дают широкие пики в спектре комбинационного рассеяния света из-за отсутствия дальнего порядка, что приводит к широкому диапазону колебательной среды для центров рассеяния.

В спектрах комбинационного рассеяния света также могут появляться комбинированные полосы, которые могут иметь различную собственную ширину по сравнению с первичными колебательными модами. Эти комбинированные полосы возникают в результате одновременного возбуждения двух или более колебательных мод и часто появляются на частотах, соответствующих сумме или разности частот отдельных мод. Они могут быть шире из-за комбинированного эффекта расширения отдельных мод.

Описанный алгоритм решает эти проблемы, используя подходы, которые допускают гибкую ширину пиков и итеративное добавление элементарных линий до тех пор, пока не будет достигнута удовлетворительная модель. Эта гибкость помогает точно моделировать широкие особенности и перекрывающиеся полосы, обычно встречающиеся в спектрах аморфных материалов и комбинированных полос.

Ошибка оптимизации модели оценивалась с помощью показателей R^2 (19), χ^2 (20), reduced χ^2 (21), Akaike Information Criterion (aic), и Bayesian Information Criterion (bic) (22).

В качестве критерия качества подгонки используется комбинация статистических показателей: коэффициент детерминации (R^2), критерий согласия (χ^2), информационный критерий Акаике (AIC) и Байесовский информационный

критерий (BIC). Это позволяет оценивать не только точность модели, но и её сложность, избегая переобучения при избыточном числе компонент.

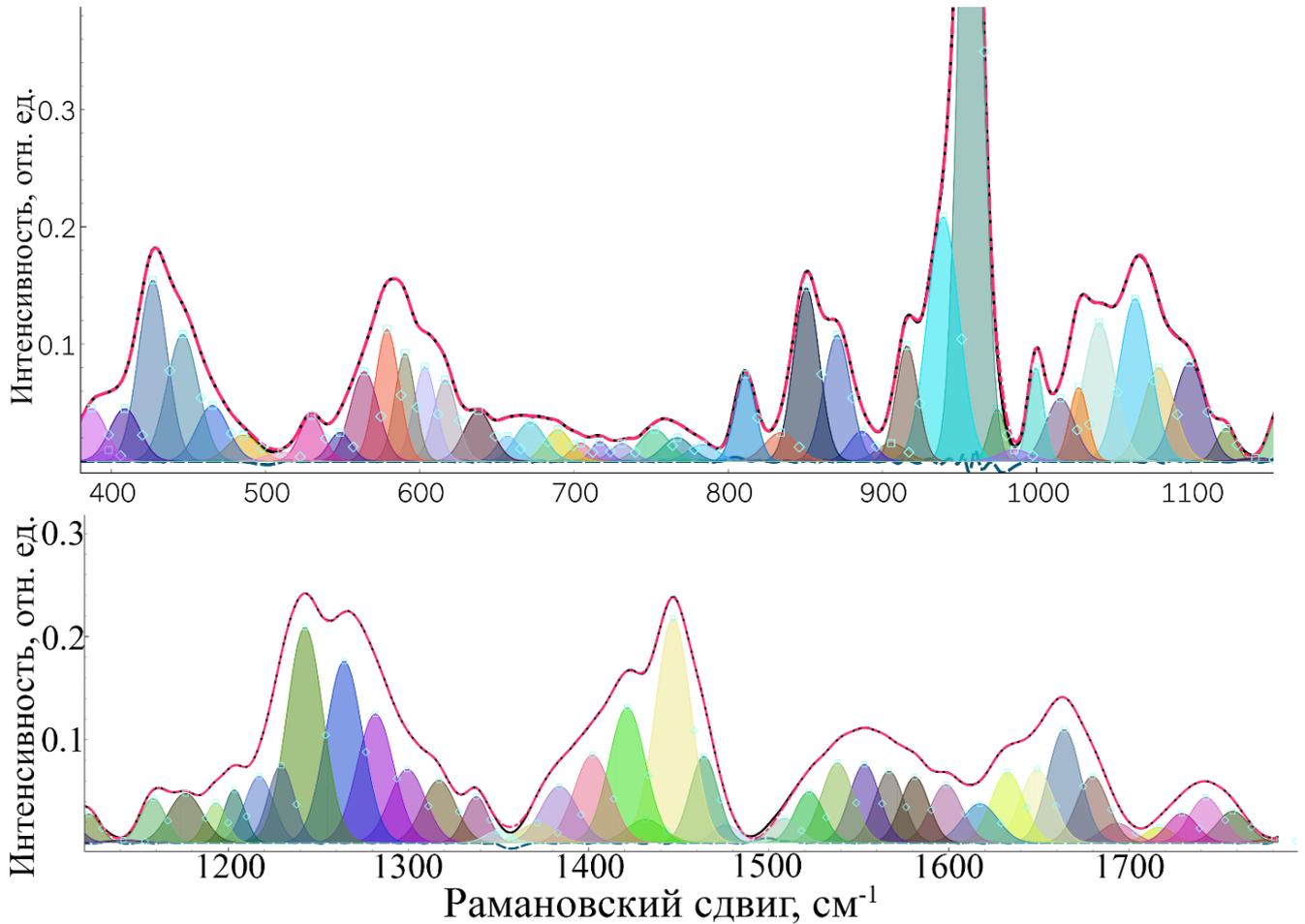


Рисунок 18 – Итоговая модель декомпозиции спектра КР. (Черная сплошная линия – исходный спектр, красная пунктирная – смоделированный спектр, синяя пунктирная – разность между исходным и смоделированным спектром)

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (19)$$

$$\chi^2 = \sum_i^N [r_i]^2, \quad (20)$$

$$\chi_v^2 = \frac{\chi^2}{N - N_{\text{varys}}}, \quad (21)$$

$$\begin{aligned} \text{aic} &= N \ln \left(\frac{\chi^2}{N} \right) + 2N_{\text{varys}} \\ \text{bic} &= N \ln \left(\frac{\chi^2}{N} \right) + \ln(N) N_{\text{varys}}, \end{aligned} \quad (22)$$

где: y – исходный спектр, f – спектр, полученный с помощью модели суммированием набора линий, g – разность между y и f , N – число точек в спектре, N_{Varys} – число параметров модели.

Значения критериев для усредненного спектра КР: $R^2 = 99,995 \%$, $\text{aic} = -1799$, $\text{bic} = -1693$, $\chi^2 = 7,3 \cdot 10^{-5}$, $\text{red } \chi^2 = 1 \cdot 10^{-6}$.

Усредненное значение этих критериев для 344 спектров КР дентина составило: $R^2 = 99,49 \%$, $\chi^2 = 1,7 \cdot 10^{-3}$, $\text{red } \chi^2 = 2,2 \cdot 10^{-5}$, $\text{aic} = -1367$, $\text{bic} = -1267$.

Оценим значение коэффициента перекрытия (overlap coefficient OVC) линий КР как отношение площади перекрытия одной линии всеми остальными к площади этой линии:

$$\text{OVC} = \frac{S_{\text{перекрытия}}}{S_{\text{одной линии}}} 100 \%, \quad (23)$$

Медианное значение коэффициента перекрытия по всем линиям в вышеуказанной модели составило 79 %.

3.4 Оценка влияния шума на качество декомпозиции

Для проверки влияния случайного гауссовского шума на результат декомпозиции спектра использовано 72 синтетических спектра КР с наложением шума. В качестве модели использован спектр с составом линий, показанным в таблице 1 и на рисунке 19.

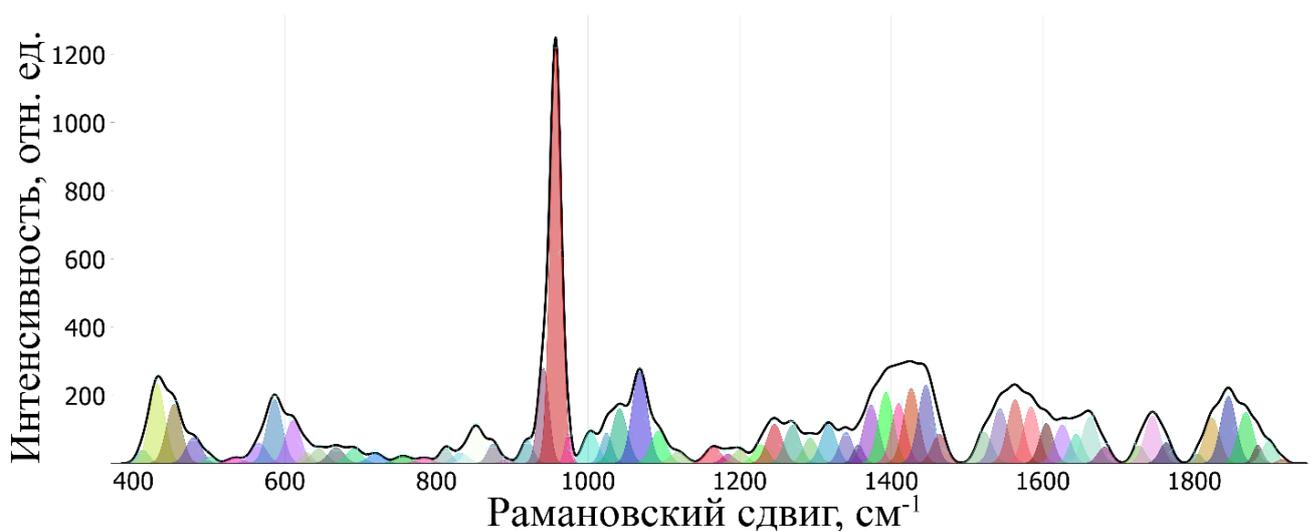


Рисунок 19 – Синтезированный спектр КР

Выборка из 72 спектров образована суммированием амплитуд линий с параметрами из таблицы 4, так что амплитуда линии изменялась в пределах $\pm 1\%$ случайным образом для моделирования выборки спектров КР костной ткани. Далее приведено сравнение результатов декомпозиции спектров КР с наложением шумов различной амплитуды для моделирования отношения сигнал/шум (ОСШ) от 1000 до 1.

На рисунке 20 показан график средней абсолютной ошибки (MAE) отклонения положения линий x_0 от действительного.

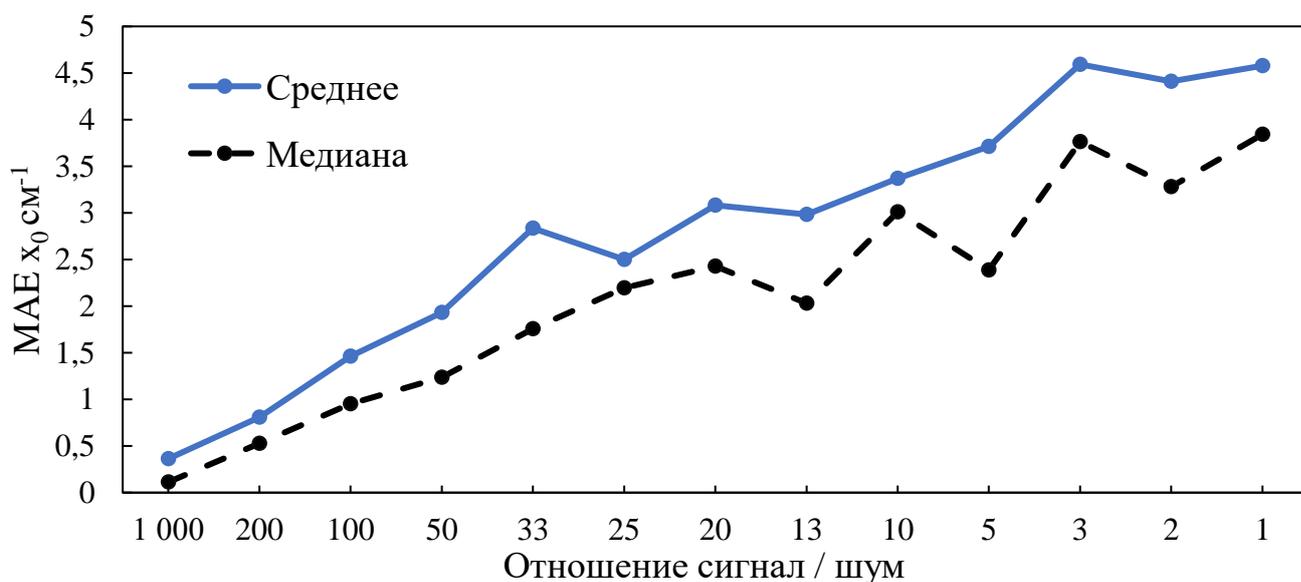


Рисунок 20 – Ошибка определения x_0 от действительного

Таблица 1. Состав линий смоделированного спектра КР.

Положение линии χ_0 , см^{-1}	Амплитуда линии a , отн. ед.	Полуширина линии НWHM $\text{d}\chi$, см^{-1}	Положение линии χ_0 , см^{-1}	Амплитуда линии a , отн. ед.	Полуширина линии НWHM $\text{d}\chi$, см^{-1}
410,0	33,19	7,93	1126,7	20,05	6,15
428,9	220,38	11,45	1158,1	27,55	8,96
449,0	173,25	12,0	1171,4	32,17	11,11
463,3	44,084	10,1	1185,5	25,27	12,0
477,7	63,85	12,0	1198,7	24,68	9,34
495,7	26,02	12,0	1210,9	20,49	11,79
532,9	17,59	9,29	1229,5	40,91	10,10
555,4	31,65	8,31	1243,2	102,92	12,0
570,1	48,27	9,56	1261,8	75,53	12,0
584,6	162,59	12,0	1274,5	63,83	11,46
596,3	57,18	10,10	1294,4	72,09	12,0
610,9	108,44	9,87	1318,2	114,30	11,99
626,8	54,11	9,64	1338,2	67,80	8,36
643,8	36,26	9,67	1355,0	83,79	12,0
663,8	48,24	12,0	1375,2	168,18	12,0
687,0	44,28	12,0	1394,1	201,50	12,0
700,4	6,75	7,14	1411,2	183,61	12,0
714,8	24,80	12,0	1427,3	198,31	12,0
731,3	12,01	10,97	1445,4	220,68	12,0
755,5	22,07	8,90	1462,5	90,82	11,78
769,8	9,48	6,35	1517,5	44,09	12,0
782,9	15,69	7,33	1531,5	100,69	12,0
791,5	6,53	5,23	1550,1	170,69	12,0
811,6	45,25	9,53	1569,0	163,12	12,0
832,3	29,80	12,0	1586,6	131,97	12,0
852,5	106,27	12,0	1605,4	114,64	12,0
874,1	55,29	8,54	1626,3	98,72	12,0
888,4	13,61	9,25	1644,1	86,93	12,0
920,8	66,77	10,25	1661,6	119,04	12,0
944,5	363,98	9,99	1679,0	54,36	12,0
957,8	1140,36	7,89	1730,4	62,78	11,72
972,2	117,11	5,90	1744,1	99,63	12,0
1003,0	95,63	11,74	1759,5	73,12	12,0
1018,4	40,40	5,56	1810,1	51,23	11,50
1026,9	72,03	6,87	1827,2	120,77	12,0
1040,9	156,71	12,0	1843,9	148,50	12,0
1066,2	233,47	12,0	1860,5	119,01	12,0
1079,7	78,49	11,25	1876,8	87,52	12,0
1095,1	67,04	9,43	1892,0	49,95	12,0
1107,8	32,20	8,34	1907,4	26,52	12,0
1119,1	19,67	6,10			

Для ОСШ ~ 1000 алгоритм определил значение x_0 со средней абсолютной ошибкой в $0,11 \text{ см}^{-1}$. С повышением уровня шума растет и ошибка определения положения линии x_0 . На рисунке 21 показана средняя абсолютная ошибка определения амплитуды линий в процентах (mean absolute percentage error MAPE).

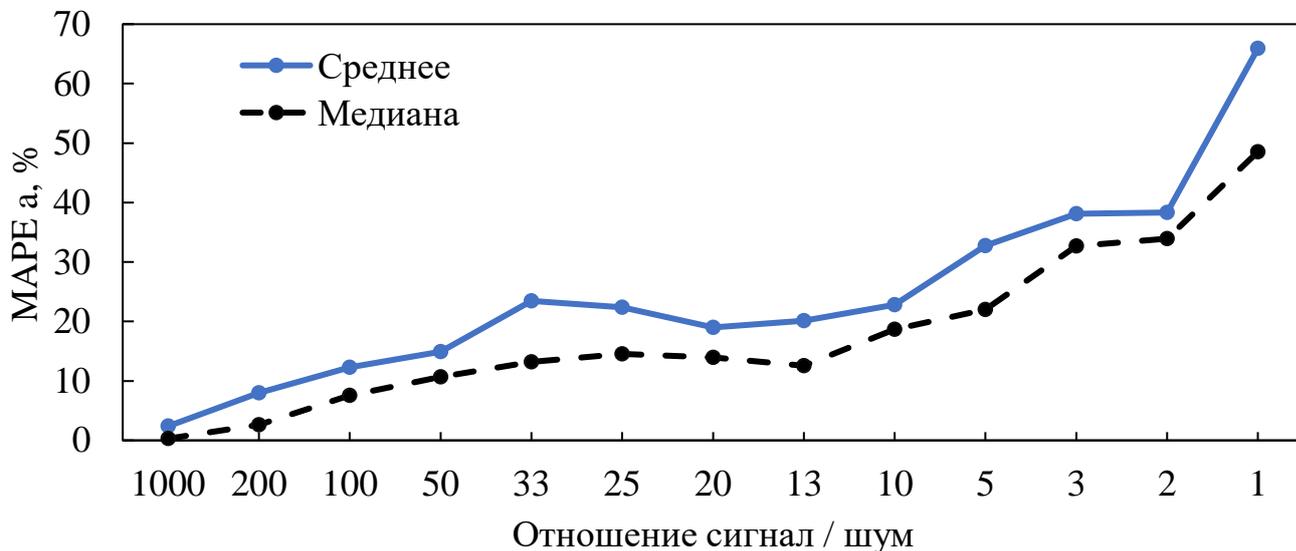


Рисунок 21 – Ошибка определения амплитуды линий

Аналогично оценивалась средняя абсолютная ошибка определения dx полуширины линий (рисунок 22).

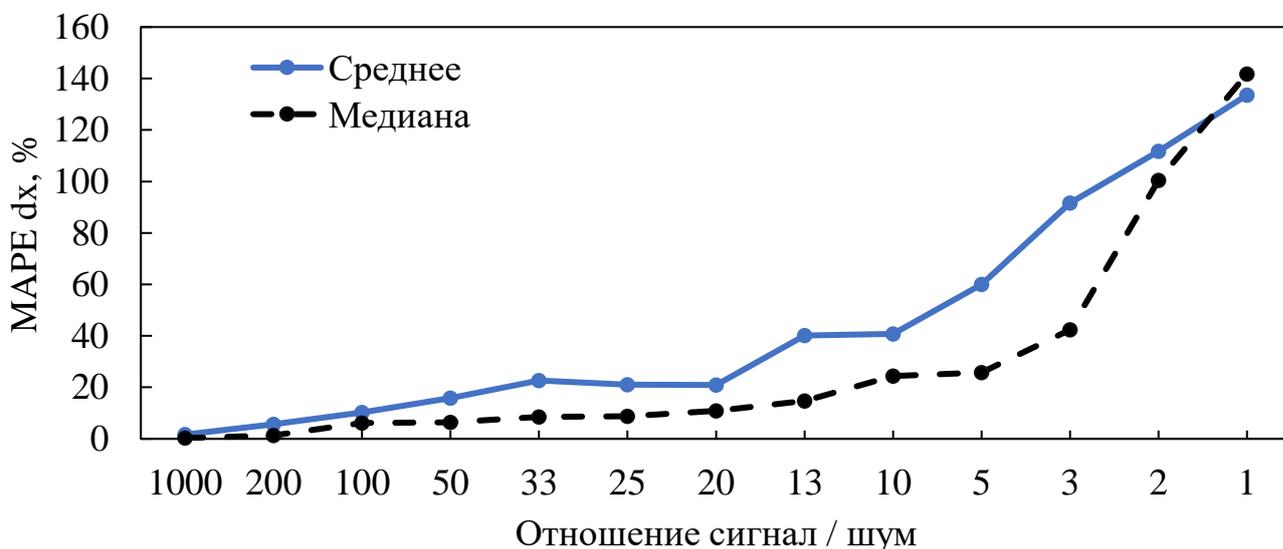


Рисунок 22 – Ошибка определения dx полуширины линий

Для уровня шума в 1 % ошибка определения амплитуды составила 0,3 %. Ошибка определения полуширины dx составила 0,29 %.

Для оценки качества алгоритма с точки зрения минимизации количества ошибочно определенных линий, сравним отношения площади ошибочно определенных линий к площади действительных линий (рисунок 23) по формуле:

$$\frac{S_{\text{ш}}}{S} * 100, \quad (24)$$

где $S_{\text{ш}}$ – площадь линий, ошибочно созданных алгоритмом, аппроксимирующих шум, S – площадь действительных линий.

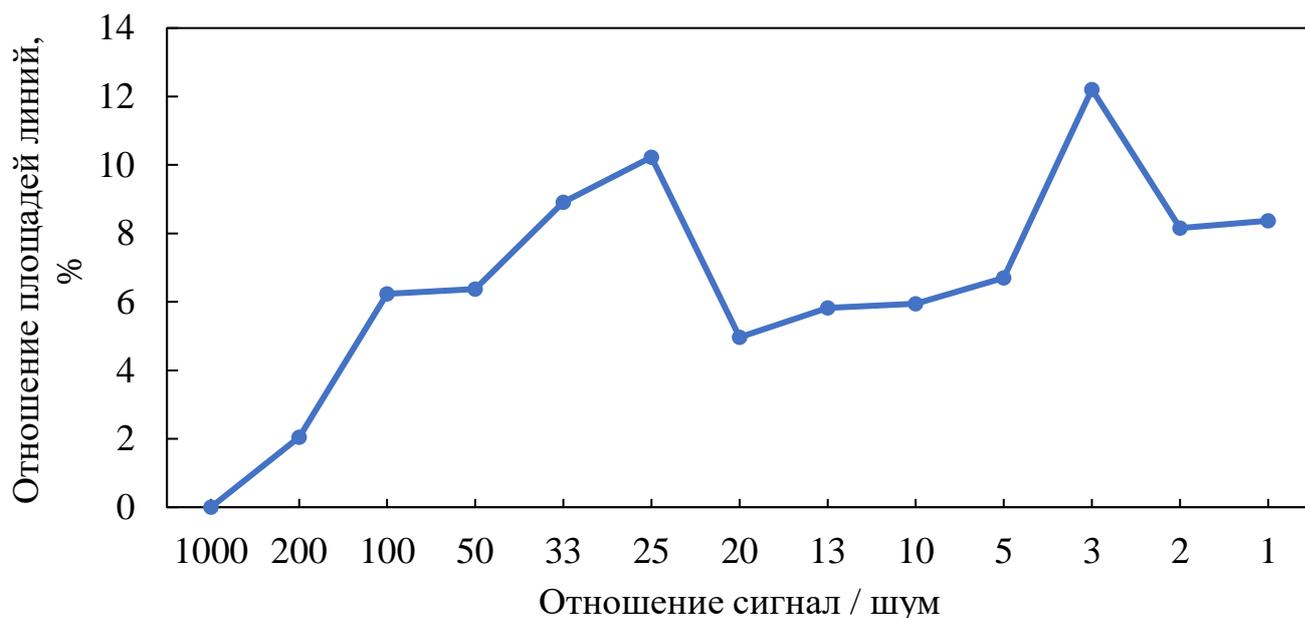


Рисунок 23 – Отношение площадей ошибочно определенных линий к действительному

Предложенный алгоритм декомпозиции линий КР позволяет получить корректный результат, соответствующий действительным параметрам модели, до отношения сигнал/шум ~ 200 . Ошибка определения амплитуды линий не превышает 15 % вплоть до ОСШ ~ 13 . Ошибка определения положения линий x_0 не превышает 1 см^{-1} до ОСШ ~ 100 , а ошибка определения полуширины линий dx не превышает 10 % вплоть до ОСШ ~ 20 .

Оценка отношения сигнал / шум на реальных данных дала результат ~ 340 ед. до процедуры сглаживания. После фильтрации данных ОСШ превышает значение в 1000 ед.

3.5 Оценка неопределенности

Доверительные интервалы часто используются на графиках, чтобы дать представление о качестве прогнозов. Найденный доверительный интервал [128] с вероятностью 0.9973 ($\sigma = 3$) для модели, описывающей спектр КР в области 988 – 1131 см^{-1} показан на рисунке 24. Доверительные интервалы (ДИ) параметров модели приведены в таблице 2.

Таблица 2. Доверительные интервалы параметров элементарных линий

$x_0, \text{см}^{-1}$	ДИ, a	ДИ, x_0	ДИ, НWHM
~1000	0,085 +/- 7,74 10^{-4} (0,91 %)	999,57 +/- 0,042	6,24 +/- 0,042(0,67 %)
~1016	0,046 +/- 0,0025 (5,41 %)	1016,27 +/- 0,67	10,99 +/- 0,021 (0,19 %)
~1030	0,11 +/- 0,0035(3,09 %)	1029,78 +/- 0,009	8,58 +/- 0,22 (2,54 %)
~1043	0,074 +/- 0,0077 (10,49 %)	1042,76 +/- 0,28	7,36 +/- 0,19 (2,53 %)
~1056	0,1 +/- 0,0068 (6,75 %)	1056,43 +/- 0,024	10,76 +/- 0,89 (8,31 %)
~1070	0,12 +/- 0,012 (9,94 %)	1070,02 +/- 0,024	10,39 +/- 0,42 (4,04 %)
~1084	0,068 +/- 0,007 (10,37 %)	1083,53 +/- 0,028	9,68 +/- 0,7 (7,27 %)
~1091	0,036 +/- 0,01 (28,92 %)	1091,21 +/- 1,14	5,83 +/- 0,64 (11,00 %)
~1099	0,051 +/- 0,023 (44,85 %)	1098,67 +/- 1,75	5,76 +/- 0,78 (13,57 %)
~1106	0,045 +/- 0,019 (42,78 %)	1106,07 +/- 2,51	6,5 +/- 1,46 (22,40 %)
~1121	0,033 +/- 0,0062 (18,78 %)	1120,96 +/- 1,31	8,96 +/- 0,95 (10,61 %)

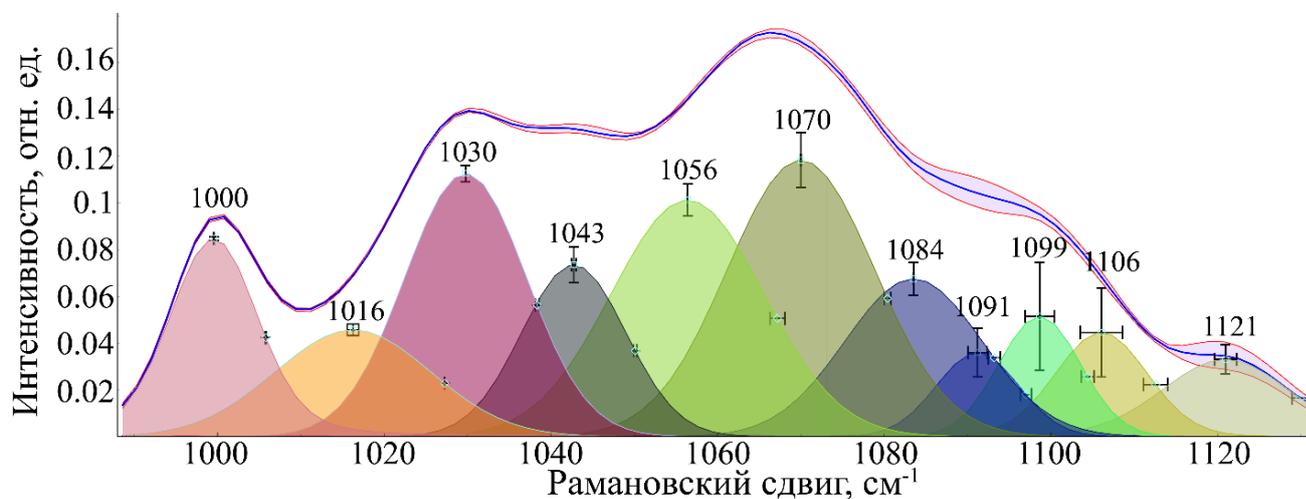


Рисунок 24 – Итоговая модель декомпозиции спектра КР в области 988 – 1131 см^{-1} . (Синяя сплошная линия – исходный спектр, красная – неопределенность)

Видно, что относительная погрешность амплитуды линий варьируется от 0,91 % до 44,85 %. Наибольшая неопределенность наблюдается в области 1080 – 1100 см^{-1} .

Оценим влияние этой неопределенности на процесс обучения модели логистической регрессии путем рандомизации амплитуды и площади линий в пределах доверительных интервалов, указанных в таблице 3.

Таблица 3. Основные метрики логистической регрессии

Данные	Точность (accuracy)	ROC AUC	Точность (Precision)	Чувствительность (Recall)	F1-score
Амплитуды	1,0	1,0	1,0	1,0	1,0
Амплитуды рандомизированные	1,0	1,0	1,0	1,0	1,0
Площадь линии	1,0	1,0	1,0	1,0	1,0
Площадь линии рандомизированная	1,0	1,0	1,0	1,0	1,0

Видно, что в пределах указанных погрешностей не происходит ухудшения метрик классификаторов, обученных на рандомизированных данных. Также не наблюдается разницы между использованием амплитуды линии и ее площади.

Далее оценим производные 1-4 порядка для обнаружения положений линий согласно критериям (13) (раздел 2.3). Согласно критериям, положение максимума линии находится в области, где совпадает отрицательный экстремум производной 2-го порядка и положительный экстремум производной 4-го порядка, при условии, что производные 1 и 3 порядка равны нулю. На рисунке 25 выделены области, удовлетворяющие этим критериям.

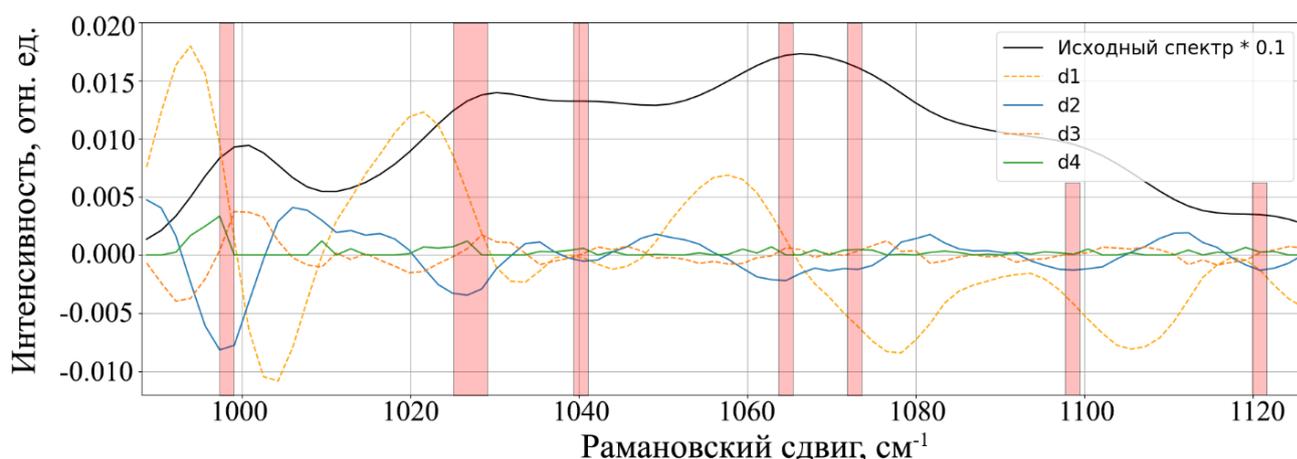


Рисунок 25 – Производные 1-4 порядка для спектра КР. Красные полосы – вероятные положения элементарных линий

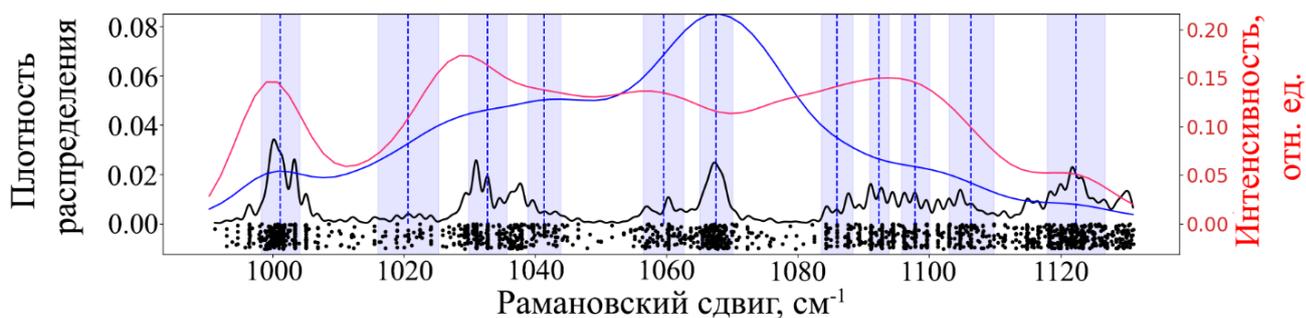


Рисунок 26 – Плотность распределения найденных координат линий (черная линия) 344 спектров КР дентина. Синяя линия – усредненный спектр минерализованного дентина, красная – деминерализованного

Положения линий примерно совпадают с результатами, определенными с помощью предложенного алгоритма декомпозиции (рисунок 26). Но параметры элементарных линии невозможно точно определить с помощью метода производных, что объясняется тем, что дифференцируется только суммарный спектр $z(\lambda)$, а непосредственно линии $z_j(\lambda)$ невозможно дифференцировать в силу их недоступности. Это порождает погрешности в определении параметров линий даже в отсутствие шумов [117].

3.6 Выводы по главе

В результате разработки был создан алгоритм декомпозиции спектров КР биологических материалов с перекрытием линий до 75 %. Этот алгоритм продемонстрировал высокую точность в определении амплитуд и положений линий в сложных спектрах биоматериалов.

Первый этап алгоритма - получение набора моделей для каждого спектрального контура. Полученные модели содержат информацию о количестве и параметрах элементарных линий, которая анализируется на втором этапе работы алгоритма. На втором этапе на основе композиции моделей с помощью кластерного анализа формируется итоговая модель, описывающая все анализируемые спектры КР.

Значения критериев качества определенной модели на тестовой выборке экспериментальных спектров КР составили: $R^2 = 99,49 \%$, $\chi^2 = 1,7 \cdot 10^{-3}$, $\text{red } \chi^2 = 2,2 \cdot 10^{-5}$, $\text{aic} = -1367$, $\text{bic} = -1267$.

Показано, что классификаторы, обученные на основании амплитуд линий декомпозированного спектра, имеют классифицирующую способность равную, классификаторам, обученных на основании полных спектров. Что говорит об отсутствии потери значимой информации при понижении размерности множества признаков и удалении линейно-зависимых признаков.

Алгоритм позволяет учитывать сдвиг положения линий КР и других параметров, работу с асимметричными линиями. Позволяет корректно выявлять линии за счет определения итоговой модели на основе композиции множества отдельных моделей.

Дополнительно была проведена оценка устойчивости алгоритма к флуктуациям входных данных. Были смоделированы шумовые спектры с $SNR = 10-50$ дБ, а также добавлены систематические искажения фона. В 94% случаев алгоритм сохранял точность восстановления положения пиков с отклонением менее $1,5 \text{ см}^{-1}$, что указывает на его высокую робастность.

Ошибка определения параметров линий КР по спектрам КР при отношении сигнал / шум = 1000, соответствующим реальным сглаженным спектрам КР составила: 0,3 % для амплитуды, 0,3 % для полуширины Δx и $0,1 \text{ см}^{-1}$ для положения линий x_0 .

На основании проведённого поиска прямого аналога предложенного алгоритма декомпозиции в виде готового программного решения не обнаружено. Большинство найденных проектов и репозиторий посвящены либо общему спектральному анализу, либо реализации отдельных алгоритмов оптимизации (например, с использованием DEAP, Inspyred или lmfit), которые требуют адаптации для задач декомпозиции спектров КР биоматериалов. Таким образом, прямого «plug-and-play» решения, полностью реализующего интегрированный алгоритм в готовом виде, в открытых источниках нет. Это позволяет считать, что данный подход уникален.

4 ПРИМЕНЕНИЕ РАЗРАБОТАННОГО СПОСОБА АНАЛИЗА СПЕКТРОВ КОМБИНАЦИОННОГО РАССЕЯНИЯ ДЛЯ ОЦЕНКИ ДЕНТИННЫХ МАТЕРИАЛОВ

Данная глава посвящена изучению возможности применения спектроскопии комбинационного рассеяния и разработанного способа обработки спектров КР для анализа дентинных материалов в процессе их деминерализации и оценки биоматериалов из ювенильного дентина и брефоматрикса.

Описано применение разработанного способа анализа спектров КР для оценки дентинных материалов. Это важно для разработки надежных способов контроля качества дентинных материалов, которые используются в стоматологии для восстановления зубов.

4.1 Применение способа спектроскопии КР для оценки дентинных материалов в процессе деминерализации

Цель исследования – апробация разработанного способа анализа спектров КР для исследования возможности оценки состава дентинных материалов в процессе их деминерализации.

4.1.1 Пробоподготовка и алгоритмы обработки спектров КР

На рисунке 27 показана блок-схема процесса пробоподготовки и регистрации спектров КР. Полученные спектры комбинационного рассеяния далее обрабатывались алгоритмами в соответствии с блок-схемой на рисунке 28 до этапа получения таблицы с амплитудами декомпозированных линий КР.



Рисунок 27 – Блок схема подготовки дентинных материалов и получения спектров КР

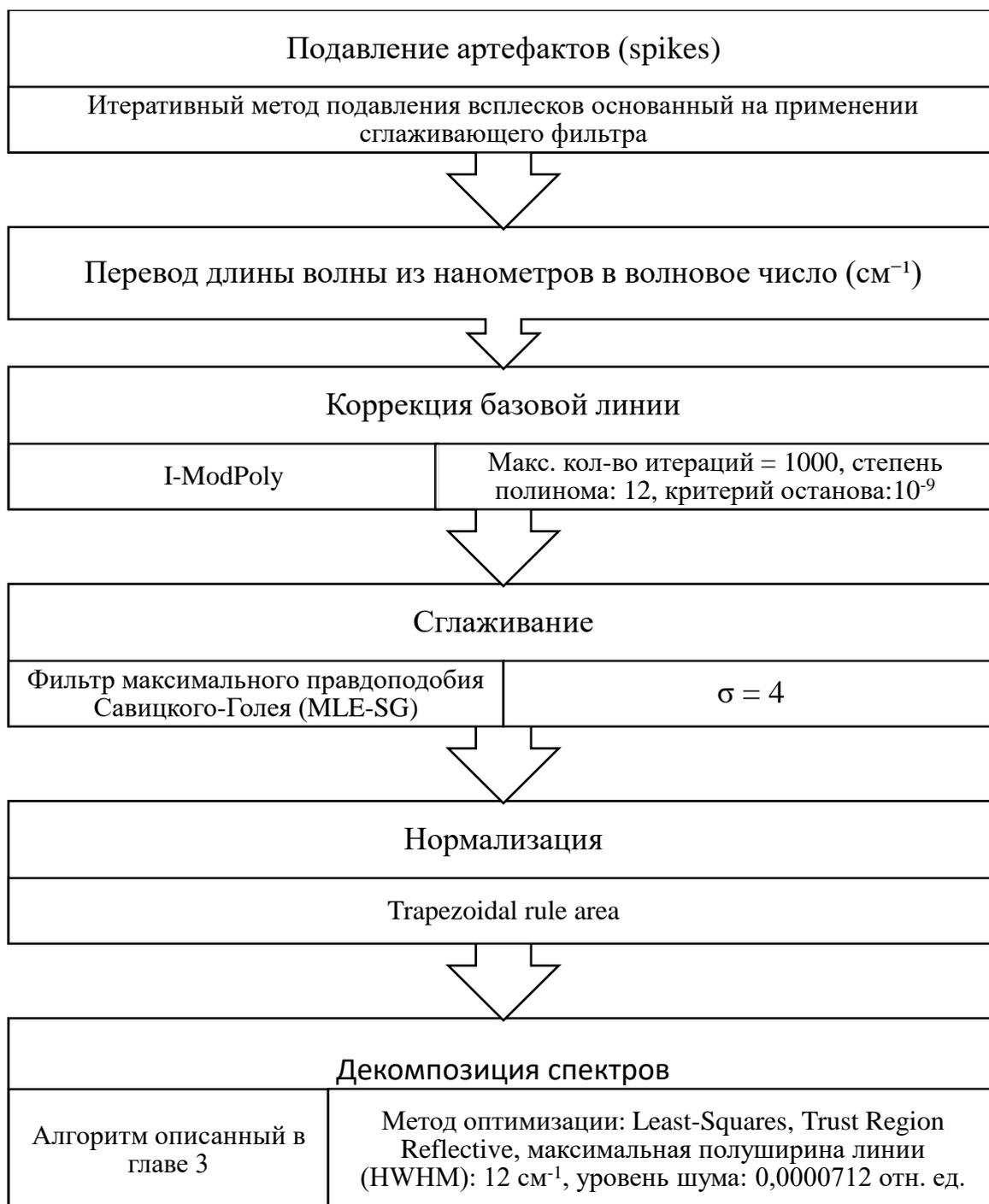


Рисунок 28 – Блок схема математической обработки спектров КР

4.1.2 Анализ усредненных спектров КР

На рисунке 29 представлены усредненные спектры по 344 спектрам КР деминерализованного и минерализованного материалов из дентина. Значительные изменения наблюдаются на линиях $850 - 875 \text{ см}^{-1}$ (C–C колебания пролина и гидроксипролина), ~ 429 , ~ 579 , $\sim 956 \text{ см}^{-1}$ (гидроксиапатит) [129], $\sim 1067 \text{ см}^{-1}$ (C–O

плоскостное валентное $\text{CO}_3^{2-} \nu_1$, $\sim 1244 \text{ см}^{-1}$ (амид III), ~ 1446 (CH_2 деформационные колебания протеина) [130] и $\sim 1664 \text{ см}^{-1}$ (амид I) [131].

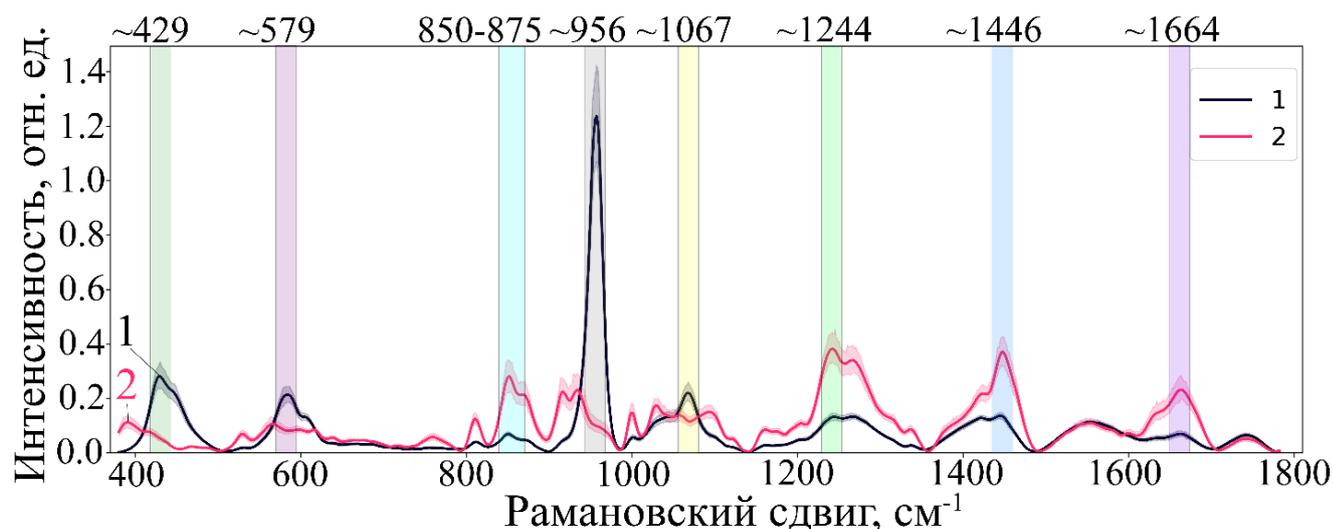


Рисунок 29 – Усредненные спектры КР дентинного материала: 1 – минерализованный; 2 – деминерализованный материал дентина

Спектральные отличия, связанные с линиями ~ 429 , ~ 579 , ~ 956 , $\sim 1067 \text{ см}^{-1}$, обусловлены вымыванием минерального компонента дентинных материалов в процессе их деминерализации.

После процесса деминерализации по технологии «ЛИОПЛАСТ» наблюдается сохранность органических составляющих, о чем свидетельствует наличие линий на $850 - 875$, ~ 1244 и $\sim 1664 \text{ см}^{-1}$, с относительной интенсивностью больше, чем в минерализованных биоматериалах.

4.1.3 Отбор признаков

Результат декомпозиции спектров КР ранее показан на рисунке 18. В результате декомпозиции были выделены 167 линии КР.

На первом этапе отбора признаков амплитуды декомпозированных линий были проанализированы методом ANOVA (Дисперсионный анализ с использованием класса SelectPercentile библиотеки sklearn с параметром percentile = 50) и из 167 были отобраны 83 признака.

На втором этапе отбора признаков был использован метод рекурсивного удаления признаков (recursive feature elimination RFE), в результате которого было отобрано 37 признаков (рисунок 30).

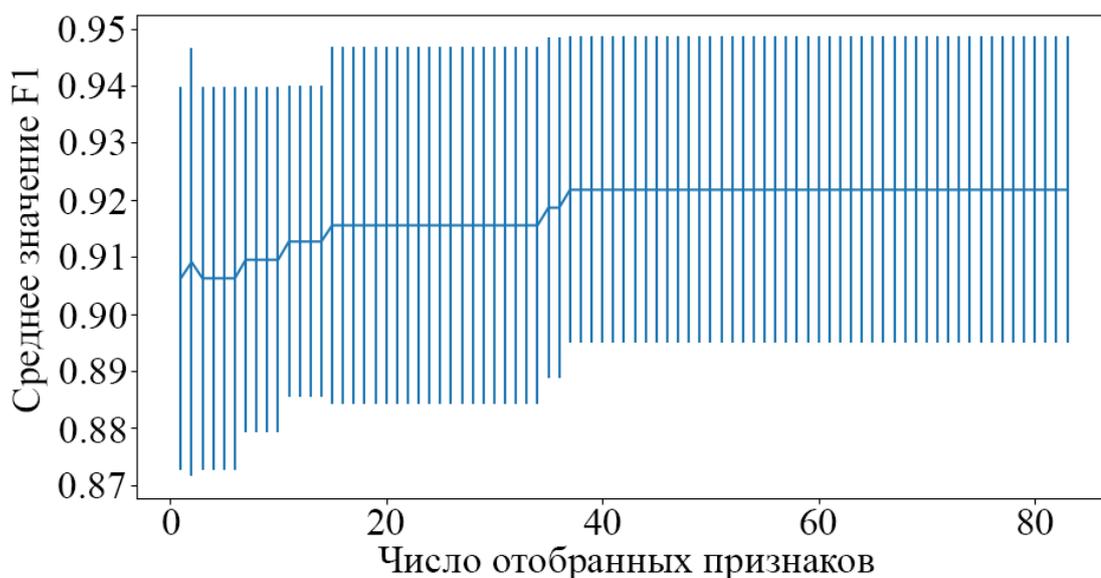


Рисунок 30 – Зависимость среднего значения метрики F1 от числа признаков в процессе рекурсивного удаления признаков

После корреляционного анализа в итоге были отобраны 5 признаков, соответствующих амплитуде декомпозированных линий на 956, 1244, 1446, 1664 и $\sim 870 \text{ см}^{-1}$.

4.1.4 Статистический анализ различий между группами

4.1.4.1 Проверка нормальности распределения

Перед проведением непараметрических тестов была проведена проверка нормальности распределения признаков в каждой из сравниваемых групп с использованием критерия Шапиро–Уилка. Полученные p -значения:

Для группы 1 (минерализованный дентин): 0,0005

Для группы 2 (демнерализованный дентин): $9,37 \cdot 10^{-6}$

Обе группы демонстрируют значимые отклонения от нормального распределения (p -значение $< 0,01$), что оправдывает использование непараметрических методов сравнения, в частности – пермутационного теста и критерия Манна–Уитни.

4.1.4.2 Пермутационный тест

Пермутационный тест (перестановочный тест, англ. permutation test) – непараметрическая процедура сравнения двух выборок, не предполагающая

нормальности распределений. Тест основан на оценке статистики различия между группами (например, расстояние между средними), а затем сравнение этого значения с распределением той же статистики, полученной в результате случайных перестановок меток группы.

Тест был реализован следующим образом:

1. Расчитано наблюдаемое расстояние между средними значениями признаков двух групп в евклидовой метрике.
2. Данные из обеих групп были объединены, затем случайным образом перемешаны метки (перестановка).
3. Для каждой перестановки пересчитывалась статистика различий.
4. Вычислено, как часто значение из перестановки превышало исходную статистику – это и есть p -значение.

Результаты теста:

- Статистика теста: 0,010;
- p -значение: 10^{-5} ;
- Расчетная мощность теста: 1,0;
- Необходимый минимальный размер выборки для достижения мощности 0,8: 5 спектров на группу.

Результаты свидетельствует о высокой чувствительности теста при текущем размере выборки и значительном эффекте между группами.

4.1.4.3 Критерий Манна–Уитни (U-тест)

Критерий Манна–Уитни – непараметрический аналог t -теста, применимый к независимым выборкам. Он оценивает, отличается ли распределение одного признака между двумя группами, без предположений о форме распределения. Этот тест особенно полезен при наличии выбросов или асимметричных распределений.

U-тест был применён к каждому признаку по отдельности. Для каждой переменной рассчитывались: статистика U и соответствующее p -значение (таблица 4).

Таблица 4. Результаты U-теста

Признак	U-статистика	<i>p</i> -значение
k870.09_a	2,0	10 ⁻⁵
k956.29_a	315,5	10 ⁻⁵
k1243.78_a	93,0	10 ⁻⁵
k1446.34_a	23,0	10 ⁻⁵
k1663.48_a	29030,5	10 ⁻⁵

Для всех признаков *p*-значение < 0,01, что указывает на значимые различия между группами по каждому из этих параметров.

Оба применённых теста (пермутационный тест и тест Манна–Уитни) показали существенные статистические различия между группами. Высокая статистическая мощность анализа подтверждает, что текущий размер выборки достаточен для обнаружения существующего эффекта.

4.1.4.4 Распределения значений признаков

На рисунке 31 и 32 представлены значения амплитуды отобранных декомпозированных линий. Видно, что для спектров минерализованного дентина характерны более высокие значения амплитуды линии ~956 см⁻¹. Для остальных линий более высокие значения характерны для спектров 2 группы образцов деминерализованного дентина. Распределение амплитуд линий 1 и 2 группы образцов позволяет хорошо различать спектры минерализованных образцов от деминерализованных.

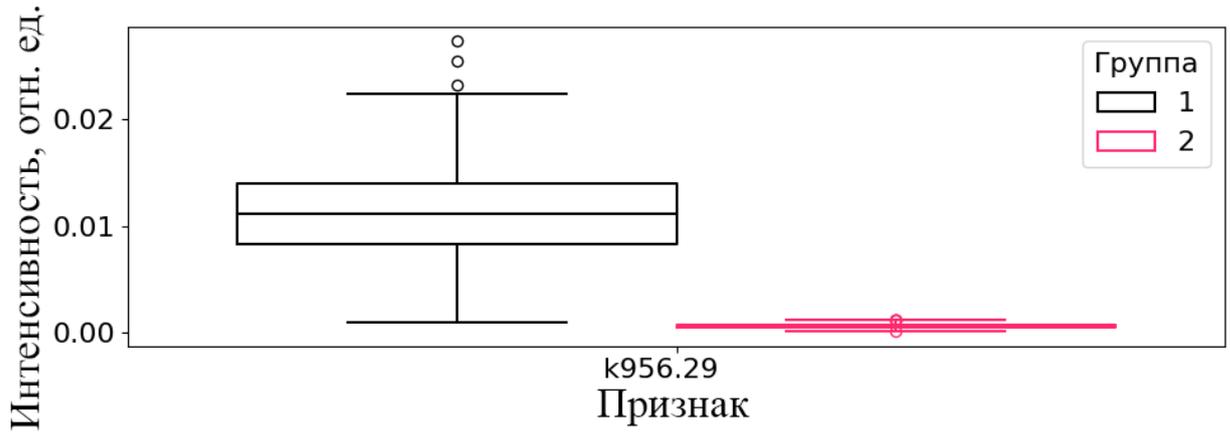


Рисунок 31 – Значения амплитуды линии $\sim 956 \text{ см}^{-1}$. 1 – минерализованный; 2 – деминерализованный материал дентина

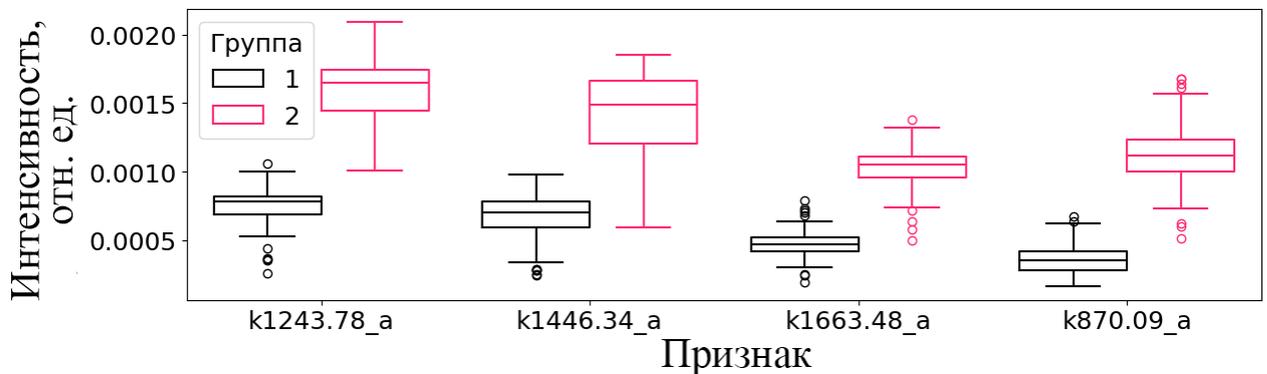


Рисунок 32 – Значения амплитуды линий ~ 870 , ~ 1244 , ~ 1446 и $\sim 1664 \text{ см}^{-1}$. 1 – минерализованный; 2 – деминерализованный материал дентина

4.1.5 Процесс создания модели машинного обучения

4.1.5.1 Выбор модели

В рамках данного исследования в качестве базовой модели машинного обучения была выбрана логистическая регрессия. Выбор данной модели обусловлен как свойствами исходных данных, так и особенностями самой модели, обеспечивающей адекватный баланс между точностью, интерпретируемостью и устойчивостью при сравнительно небольшой размерности пространства признаков.

Анализ структуры выборки показал, что:

- Данные представлены в виде числовых векторов;
- После отбора осталось всего пять признаков;

- Наблюдается линейная зависимость между признаками и целевой переменной;
- Объём выборки ограничен, что требует использования моделей с низким риском переобучения;
- Исходя из статистического анализа распределений значений признаков и гипотезы, что существует линейная разделяющая граница, возможно применение логистической регрессии.

Логистическая регрессия представляет собой линейную классификационную модель, которая прогнозирует вероятность принадлежности объекта к одному из классов. Основные аргументы в пользу её выбора:

- Интерпретируемость: модель позволяет анализировать влияние каждого признака на итоговое решение (веса модели);
- Устойчивость при ограниченной размерности: не требует большого количества данных для обучения;
- Вероятностный выход: позволяет получать вероятности классов без необходимости дополнительной калибровки;
- Быстрота обучения: особенно важна при множественных запусках (например, в кросс-валидации);
- Наличие калиброванного выхода: логистическая регрессия изначально выдаёт вероятности, согласованные с истинными (при выполнении условий линейности).

Тем не менее, следует учитывать и недостатки модели:

- Необходимость масштабирования признаков: логистическая регрессия чувствительна к шкале входных данных;
- Линейность границы разделения: не способна корректно описывать сложные нелинейные зависимости без трансформации признаков;
- Чувствительность к выбросам при отсутствии регуляризации.

Таким образом, логистическая регрессия в данном контексте является наиболее обоснованным выбором, обеспечивая баланс между

интерпретируемостью, простотой и достаточной точностью при ограниченном объёме и размерности данных.

Важным критерием выбора логистической регрессии стало то, что она изначально обучается с вероятностной функцией потерь (логистическая функция). В результате, предсказанный скор (вероятность) отражает степень уверенности модели в принадлежности объекта к классу, и в отличие от многих других моделей, эти значения уже калиброваны.

Это особенно важно в контексте данной работы, поскольку одной из задач является оценка деминерализации. Промежуточные значения вероятности (например, 0,5, 0,6, 0,7 и т.д.) можно интерпретировать как непрерывную шкалу, отражающую прогрессирование процесса деминерализации, а не только бинарное решение.

4.1.5.2 Подготовка выборки для обучения модели логистической регрессии

Для построения модели логистической регрессии использовались данные, представляющие таблицу, в которой каждый спектр характеризуется пятью признаками, отражающими амплитуду декомпозированных линий комбинационного рассеяния: 956, 1244, 1446, 1664 и ~ 870 см^{-1} .

Образцы распределены по двум группам:

- Группа 1 (минерализованный материал): 194 спектра,
- Группа 2 (деминерализованный материал из дентина): 150 спектров.

Несмотря на некоторое числовое различие между группами, дисбаланс классов является незначительным, что обеспечивает корректное представление распределения классов как в обучающей, так и в тестовой выборках.

Выборка разделялась на обучающую и тестовую с помощью функции `train_test_split` из библиотеки `scikit-learn`. При этом использовались следующие параметры:

- `test_size` задаётся через значение, полученное из пользовательского интерфейса (например, `test_data_ratio_spinBox.value() / 100.`), что позволяет гибко настраивать процент от исходной выборки, выделяемый под тестовые данные.

- `stratify=y` обеспечивает пропорциональное распределение классов в обеих выборках, что особенно важно при наличии даже небольшого дисбаланса между группами.
- `random_state` фиксирует случайность процесса разделения, что гарантирует воспроизводимость эксперимента.

Пример кода:

```
x_train, x_test, y_train, y_test = train_test_split(
    x, y,
    test_size=mw.ui.test_data_ratio_spinBox.value() / 100.,
    stratify=y, random_state=rnd_state)
```

В данном исследовании параметр `test_size = 25`, что соответствует размеру тестовой выборки 25 % от общего числа наблюдений (344).

Перед обучением модели данные подвергаются дополнительной обработке, направленной на улучшение качества обучения. В рамках конвейера обработки используется `QuantileTransformer` с параметром `output_distribution='normal'`.

`QuantileTransformer` – это метод нелинейного масштабирования, который преобразует распределение признаков так, чтобы оно соответствовало заданному типу распределения, в нашем случае – нормальному. Преимущества данного подхода заключаются в следующем:

- Стабилизация распределения признаков: Признаки, имеющие сильно скошенные или несимметричные распределения, преобразуются таким образом, что их эмпирическая функция распределения становится приблизительно нормальной. Это способствует более стабильной и быстрой сходимости модели.
- За счёт применения квантильного масштабирования экстремальные значения «сжимаются», что снижает их влияние на модель.
- Благодаря применению `QuantileTransformer` модель логистической регрессии оказалась откалибрована, что подтверждается корректной интерпретацией вероятностных прогнозов. В отличие от стандартного масштабирования (например, `StandardScaler`), `QuantileTransformer` не ограничивается лишь смещением и масштабированием по параметрам среднего и

стандартного отклонения, а учитывает всю форму эмпирического распределения, что может приводить к более качественному приведению данных к требуемой нормальной форме.

4.1.5.3 Обучение модели логистической регрессии

Для построения базовой модели классификации использовалась логистическая регрессия, включённая в состав конвейера обработки данных (Pipeline), что обеспечивает последовательное применение всех преобразований как к обучающей, так и к тестовой выборкам без утечки информации, вместе с предварительным масштабированием признаков с помощью `QuantileTransformer` (`output_distribution='normal'`) и калибровки модели с помощью сигмоиды. Обучение модели производилось без дополнительного подбора гиперпараметров, с использованием параметров по умолчанию, которые в данном случае являются обоснованным выбором (таблица 5). Конфигурация модели приведена ниже:

```
lr = LogisticRegression(class_weight='balanced', max_iter=1000)
model = Pipeline(['scaler',
QuantileTransformer(output_distribution='normal')), ('clf', lr)]
calibrated_model = CalibratedClassifierCV(model, method='sigmoid',
cv=StratifiedKFold(n_splits=10, shuffle=True, random_state=rnd))
```

Таблица 5. Ключевые параметры и их значение по умолчанию.

Параметр	Значение	Обоснование
penalty	'l2'	L2-регуляризация сглаживает веса, уменьшая риск переобучения
C	1,0	Обратная сила регуляризации. Значение 1.0 обеспечивает умеренный уровень штрафа
solver	'lbfgs'	Быстрый и устойчивый оптимизатор, подходящий для задач с малым числом признаков и двумя классами
max_iter	1000	Увеличенное количество итераций обеспечивает сходимость при трансформированных данных
class_weight	'balanced'	Автоматически компенсирует незначительный дисбаланс классов

Такой подход гарантирует, что:

- Признаки из обучающей выборки преобразуются с использованием `QuantileTransformer`, и те же параметры преобразования применяются к тестовой выборке.
- Логистическая регрессия обучается на данных, приведённых к нормальному распределению, что положительно сказывается на стабильности и точности модели.
- Выход модели содержит откалиброванные вероятности принадлежности к классу 2 (деминерализованная группа).

В `scikit-learn` логистическая регрессия по умолчанию использует L2-регуляризацию, которая помогает предотвратить переобучение путём штрафа за слишком большие значения коэффициентов модели. L2-регуляризация является особенно эффективной при наличии высокоррелированных признаков и работает хорошо при ограниченном количестве признаков, как в данной задаче (всего 5 признаков).

Кроме того, выбранная стратегия обучения не использует гиперпараметрический поиск (например, через `Optuna` или `GridSearchCV`), что обусловлено следующими факторами:

- Базовая модель уже демонстрирует высокое качество классификации.
- Простота модели позволяет обеспечить интерпретируемость коэффициентов и прогнозов.
- Отсутствие необходимости в тщательной настройке снижает риск переобучения.

При построении базовой (baseline) модели логистической регрессии в данной работе кросс-валидация целенаправленно не использовалась, что обусловлено несколькими факторами:

- В исследовании используются всего пять признаков, что делает модель менее подверженной переобучению. Логистическая регрессия, как линейная

модель, хорошо работает на таких относительно низкоразмерных данных, особенно если между признаками отсутствует мультиколлинеарность.

- Разделение выборки на обучающую и тестовую происходило с использованием стратификации по целевому признаку. Это обеспечивает сопоставимое распределение классов в обеих подвыборках. Кроме того, дисбаланс между классами практически отсутствует (194 и 150 наблюдений), что дополнительно снижает необходимость в регулярной переоценке модели на разных фолдах.

- Оценка качества производится на отложенной тестовой выборке. Поскольку данные были заранее разделены на обучающую и тестовую подвыборки, модель может быть оценена по результатам на тестовой части, которая не использовалась при обучении. Это позволяет получить независимую оценку обобщающей способности модели без необходимости во внутренней перекрёстной проверке.

4.1.6 Оценка качества и интерпретация полученного классификатора

В таблице 6 приведены значения метрик модели логистической регрессии. Модель демонстрирует высокие значения метрик как на обучающей, так и на тестовой выборках. Подобные результаты могут свидетельствовать о полной линейной разделимости классов в пространстве признаков после трансформации, а также об отсутствии переобучения, что подтверждается идентичными метриками на тренировочной и тестовой выборках.

В таблице 7 показаны коэффициенты модели. Признак k956.29_a имеет отрицательное влияние на вероятность положительного класса, тогда как остальные признаки – положительное. Это свидетельствует о потенциальной анти-корреляции между k956.29_a и целевой переменной.

Таблица 6. Значения метрик модели логистической регрессии, доверительные интервалы рассчитаны для уровня вероятности не менее 95 %

Метрика	Обучающая выборка (258 спектров)	Тестовая выборка (86 спектров)
Точность (accuracy)	1,0 (1,0-1,0)	1,0 (1,0-1,0)
ROC AUC	1,0 (0,99-1,0)	1,0 (0,99-1,0)
Точность (precision)	1,0 (1,0-1,0)	1,0 (1,0-1,0)
Чувствительность (recall)	1,0 (1,0-1,0)	1,0 (1,0-1,0)
Специфичность	1,0 (1,0-1,0)	1,0 (1,0-1,0)
Оценка F1 (F-мера)	1,0 (1,0-1,0)	1,0 (1,0-1,0)
Логарифмическая потеря (Logloss)	0,097 (0,088-0,11)	0,080 (0,067-0,094)

Таблица 7. Коэффициенты признаков модели

Признак	Коэффициент
k956.29_a	-1,971508
k1446.34_a	1,285561
k1243.78_a	1,188034
k870.09_a	1,174798
k1663.48_a	1,162568

Распределение предсказанных значений (вероятность принадлежности к классу 2) (рисунок 33) демонстрирует значительное смещение средних значений и указывает на способность модели учитывать различия между классами.

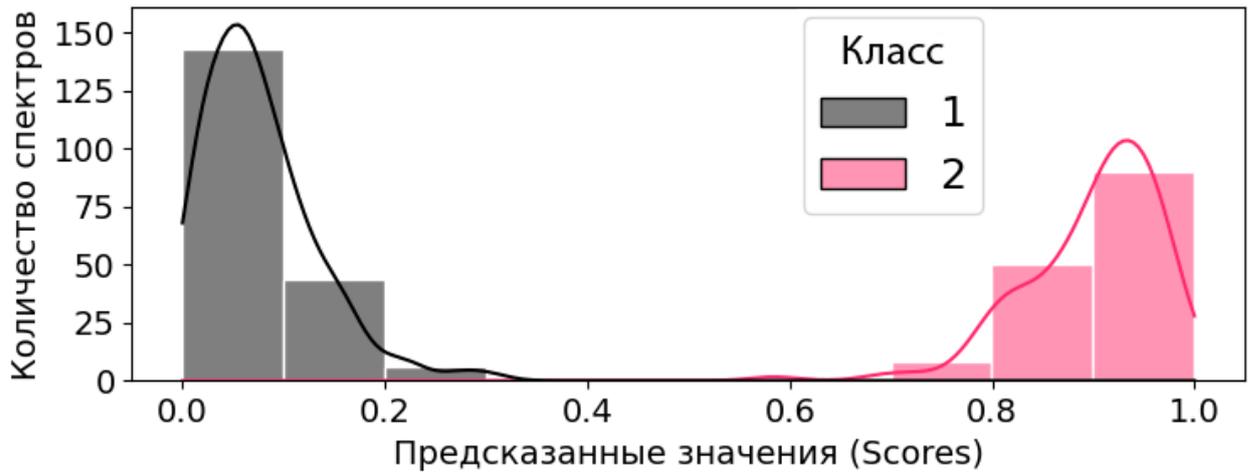


Рисунок 33 – Значения функции принятия решения (decision score). 1 – минерализованный; 2 – деминерализованный материал дентина

На проекции с использованием МГК (рисунок 34) наблюдается четкое разделение классов. Это подтверждает гипотезу о линейной разделимости в исходных или трансформированных данных. Этот график представляет собой визуализацию решающих границ и классификационных результатов модели логистической регрессии в двумерном пространстве, полученном с помощью МГК.

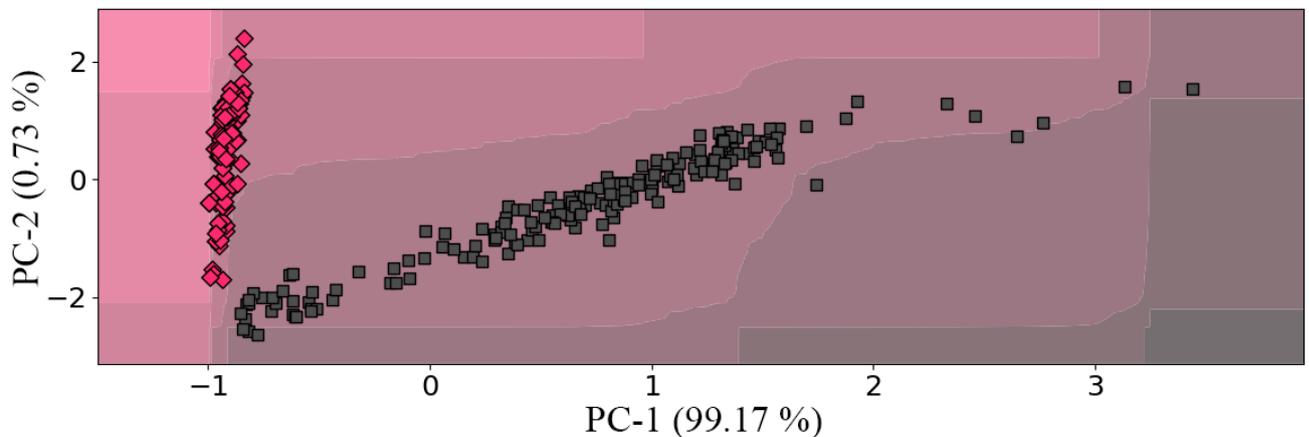


Рисунок 34 – Проекция решений модели для данных в пониженной размерности.

Черный – минерализованный дентин;

Красный – деминерализованный материал дентина

PC-1 объясняет практически всю дисперсию (99,17 %), что означает, что данные практически одномерны. Это может указывать на то, что вторая компонента (PC-2) малоинформативна, и модель в основном использует PC-1 для классификации, соответствующий признаку k956.29_a.

Матрица ошибок (рисунок 35) показывает отсутствие ошибок классификации: все объекты классифицированы корректно, как в положительном, так и в отрицательном классах.

Истинный класс	1	48	0
	2	0	38
		1	2
		Предсказанный класс	

Рисунок 35 – Матрица решений

Кривая ошибок первого и второго рода (ROC AUC curve) (рисунок 36) поднимается строго к верхнему левому углу, площадь под кривой (AUC) составляет 1, что подтверждает абсолютную способность модели различать классы безошибочно. Анализ кривой точности – полноты (Precision-Recall curve) (рисунок 36) показал, что логистическая регрессия достигает значения средней точности (Average Precision) ($AP = 1,0$), что свидетельствует о полном отсутствии ошибок классификации. Это может объясняться высокой степенью делимости классов в пространстве признаков.

Для интерпретации вклада отдельных признаков в предсказания модели логистической регрессии был построен график типа SHAP beeswarm plot (рисунок 37), отражающий влияние каждого признака на выход модели. SHAP (SHapley Additive exPlanations) [132] – это метод объяснения работы модели, основанный на теории кооперативных игр Шепли. Он позволяет количественно оценить, насколько каждый признак влияет на итоговое предсказание по отношению к базовому значению (intercept).

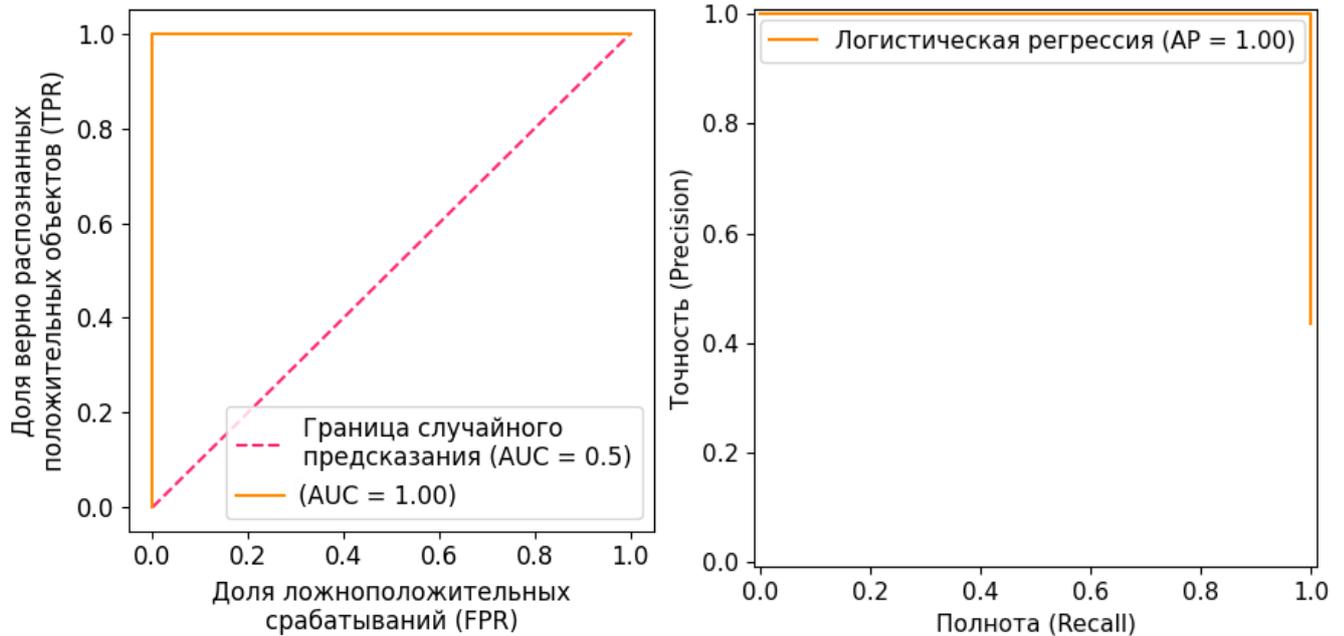


Рисунок 36 – Кривая ошибок первого и второго рода (ROC AUC curve) (слева) и кривая точности – полноты (Precision – Recall curve) (справа)

На графике по вертикальной оси отображены признаки, отсортированные по убыванию их важности. По горизонтальной оси показано значение SHAP, характеризующее величину и направление вклада соответствующего признака в выход модели. Цвет каждой точки отражает значение признака: от синего (низкое значение) до красного (высокое значение). Таким образом, каждая точка на графике соответствует одному примеру из обучающей выборки.

Признак k956.29_a оказывает наибольшее влияние на модель. При низких значениях (синие точки) SHAP-значения положительные, то есть модель склонна относить такие образцы к классу 2 (деминерализованный дентин). При этом более высокие значения этого признака снижают вероятность отнесения к классу 2, и, соответственно, модель классифицирует такие образцы как минерализованные (класс 1). Это указывает на инвертированный характер влияния этого признака.

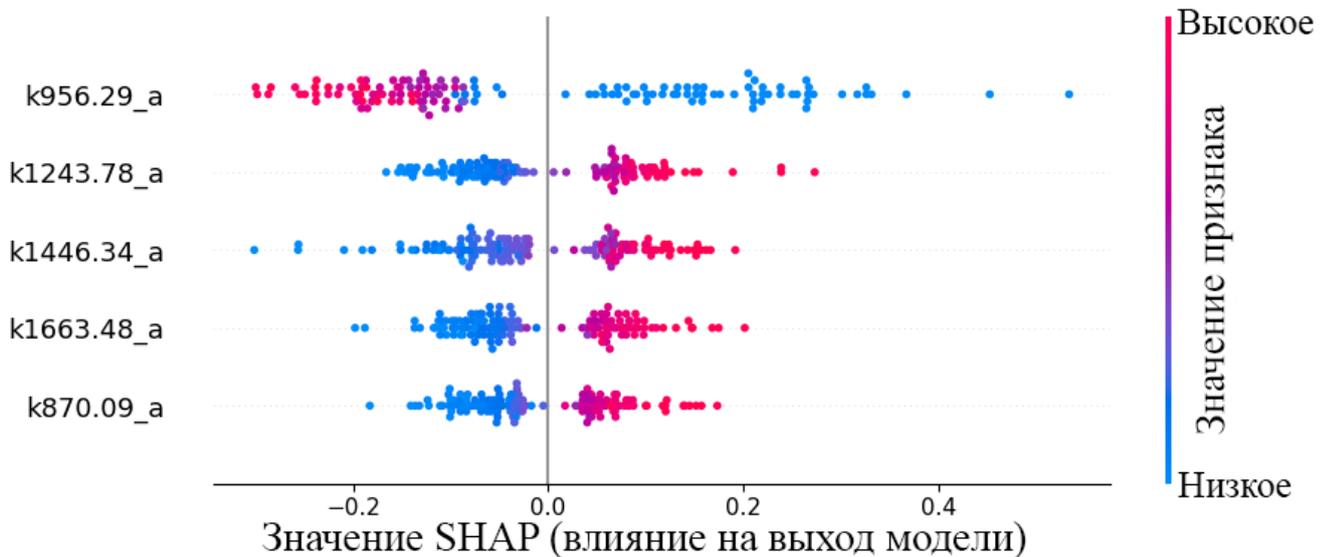


Рисунок 37 – график SHAP-значений для логистической регрессии. По оси Y отложены линии КР, по оси X – величина их вклада в предсказание модели (SHAP). Красный цвет соответствует высоким значениям признака, синий – низким. Положительные SHAP-значения увеличивают вероятность целевого класса (деминерализованный дентин), отрицательные – уменьшают

Для остальных признаков (например, k1446.34_a, k1243.78_a, k1663.48_a, k870.09_a) наблюдается противоположная зависимость: более высокие значения (красные точки) связаны с положительными SHAP-значениями, а значит, способствуют классификации в пользу класса 1 – деминерализованного дентина.

Суммарно, видно, что деминерализованный дентин характеризуется высокими значениями большинства признаков, за исключением k956.29_a, где наблюдается снижение.

Минерализованный дентин (класс 1), наоборот, характеризуется высокими значениями только признака k956.29_a, а по остальным признакам значения преимущественно низкие.

Рисунок 38 представляет собой тепловую карту (heatmap) SHAP-значений, которая иллюстрирует вклад каждой спектральной линии (признака) в итоговое предсказание модели. По оси Y расположены линии КР, отсортированные по их важности, а по оси X – отдельные наблюдения (спектры). Цветовая гамма используется для обозначения направления вклада:

- Красный цвет указывает на то, что данный признак увеличивает вероятность отнесения объекта к целевому классу (в данном случае – деминерализованному дентину),
- Синий цвет означает, что признак снижает эту вероятность (то есть способствует классификации в группу минерализованного дентина).

Функция $f(x)$, указанная в верхней части рисунка, представляет собой "сырые оценки" или logits, то есть значения линейной комбинации признаков до применения сигмоидальной функции. Эти значения отражают уверенность модели в предсказании:

- Высокое положительное значение $f(x)$ указывает на сильное доверие модели к принадлежности объекта к классу с деминерализованным дентином,
- Высокое отрицательное значение – к минерализованному дентину.

Тепловая карта показывает, что признаки существенно различаются между группами. Для группы минерализованного дентина наблюдаются высокие значения лишь по признаку k956.29_a (синяя доминанта), в то время как для деминерализованного дентина большинство других признаков (k1446.34_a, k1243.78_a, k1663.48_a, k870.09_a) имеют высокие значения, отраженные красным цветом. Это свидетельствует о том, что модель эффективно различает классы благодаря значимым различиям в спектральных характеристиках.

Несмотря на высокие результаты на обучающей и тестовой выборках, высокая точность может быть связана с полным разделением классов в рамках текущей выборки. Поэтому для окончательной оценки качества модели крайне важно провести валидацию на независимой выборке, что позволит проверить устойчивость и обобщающую способность модели.

Использование SHAP-анализа позволило наглядно интерпретировать вклад каждого спектрального признака в предсказания модели. Калиброванные выходные значения логистической регрессии обеспечивают возможность интерпретации промежуточных значений скор, что позволяет оценивать не только бинарное решение. Модель выдаёт скор [100; 0] для минерализованного дентина (группа 1) и [0; 100] для деминерализованного (группа 2).

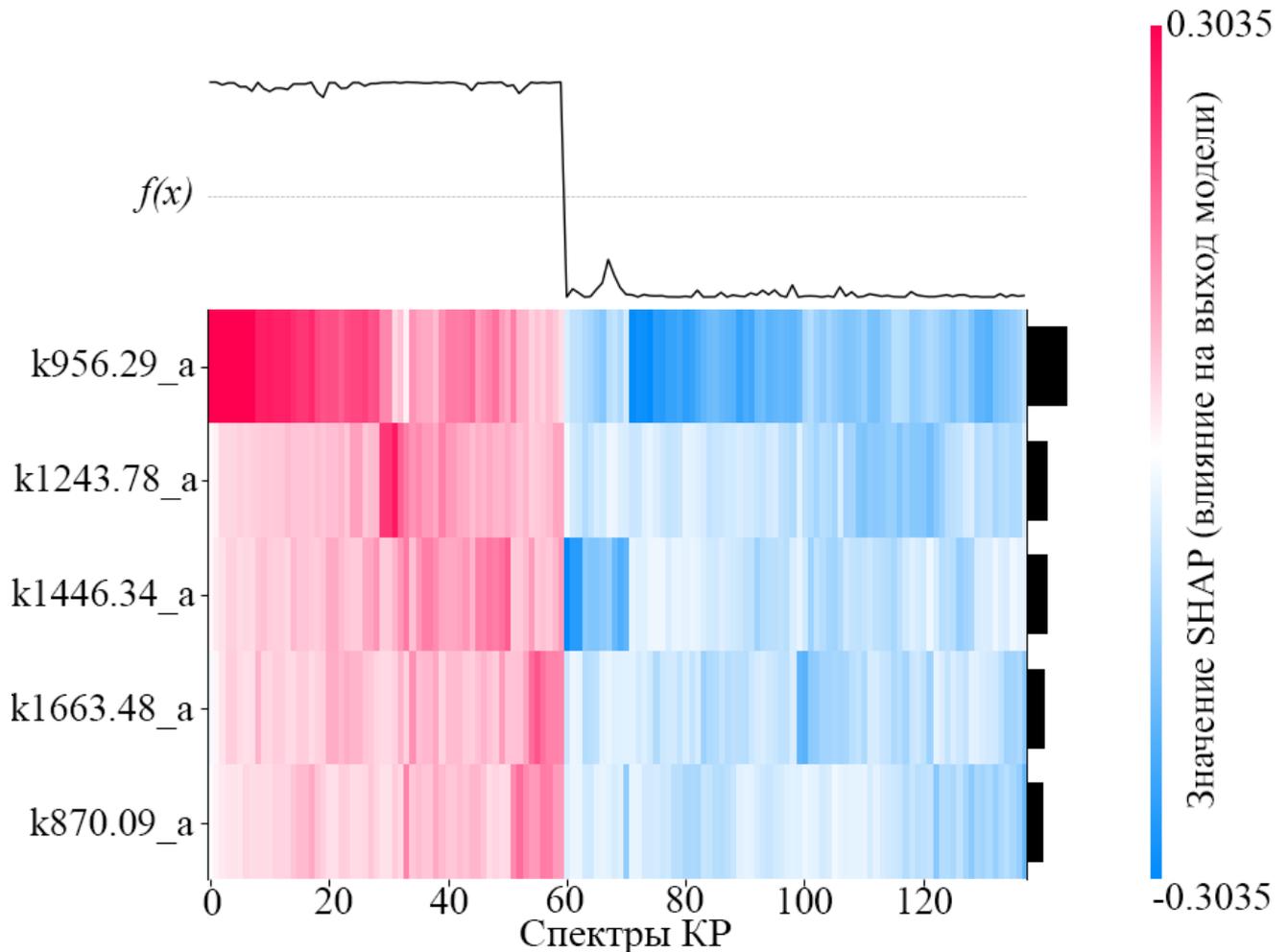


Рисунок 38 – Тепловая карта SHAP-значений. По оси Y отложены линии КР, упорядоченные по важности, по оси X – отдельные наблюдения (спектры). Красный цвет указывает на увеличение вероятности целевого класса, синий – на уменьшение. Значения SHAP нормированы в диапазоне от $-0,3035$ до $0,3035$.

В качестве показателя коррелирующего с фактом деминерализации используется decision score логистической регрессии (диапазон 0–100 баллов), где 0 соответствует полностью минерализованному материалу, а 100 – полностью деминерализованному. Дополнительно оценка основывается на пяти амплитудах декомпозированных линий КР (см^{-1}): k_{956} , k_{870} , k_{1244} , k_{1446} , k_{1664} .

Критерии завершённой деминерализации:

Для каждого из пяти признаков вычисляется эмпирический квантиль q (от 0 до 1).

1. Лёгкая деминерализация (decision score 0–33):

- амплитуда линии $k_{956} \text{ см}^{-1}$ снижается не более чем на 20 % относительно контрольной (минерализованной) группы;
 - остальные четыре амплитуды (k_{870} , k_{1244} , k_{1446} , k_{1664}) превышают значения минерализованного дентина не более чем на 30 %.
 - $q_i \geq 0,67$ для большинства признаков (то есть амплитуда близка к верхним 33 % значений «здорового» дентина).
2. Умеренная деминерализация (decision score 34–66):
 - амплитуда $k_{956} \text{ см}^{-1}$ снижается на 20–50 %;
 - амплитуды k_{870} , k_{1244} , k_{1446} , k_{1664} повышены на 30–70 %.
 - $0,33 < q_i < 0,67$ (амплитуды лежат в среднем интервале распределения).
 3. Выраженная деминерализация (decision score 67–100):
 - амплитуда $k_{956} \text{ см}^{-1}$ снижается более чем на 50 %;
 - амплитуды остальных линий превышают минерализованный уровень более чем на 70 %.
 - $0,33 \leq q_i$ (амплитуды близки к нижним 33 % – существенно снижены).

4.2 Сравнительная характеристика биоматериалов из ювенильного дентина и брефоматрикса

В данном разделе приведен сравнительный анализ спектральных свойств новых биоматериалов из дентина ювенильных зубов и брефоматрикса из кортикальной и губчатой костной ткани.

Целью исследования являлась апробация разработанного способа при проведении сравнительного анализа спектров КР ювенильного дентина и брефоматрикса из кортикальной и губчатой костной ткани.

4.2.1 Пробоподготовка и алгоритмы обработки спектров КР

На рисунке 39 показана блок-схема процесса пробоподготовки и регистрации спектров КР. Полученные спектры комбинационного рассеяния далее

обрабатывались алгоритмами в соответствии с блок-схемой на рисунке 40 до этапа получения таблицы с амплитудами декомпозированных линий КР.



Рисунок 39 – Блок схема подготовки образцов и получения спектров КР



Рисунок 40 – Блок схема математической обработки спектров КР

4.2.2 Анализ усредненных спектров КР

На рисунке 41 представлены нормированные спектры КР исследуемых образцов. Спектры комбинационного рассеяния всех групп имели спектральные различия. Также видно, что спектры брeфоматрикса из кортикальной костной ткани (вторая группа) ближе к спектрам биоматериалов, приготовленных из ювенильного дентина. Вероятно, это связано со сходством состава плотной

кортикальной кости брефоматрикса с ювенильными зубами после деминерализации [133].

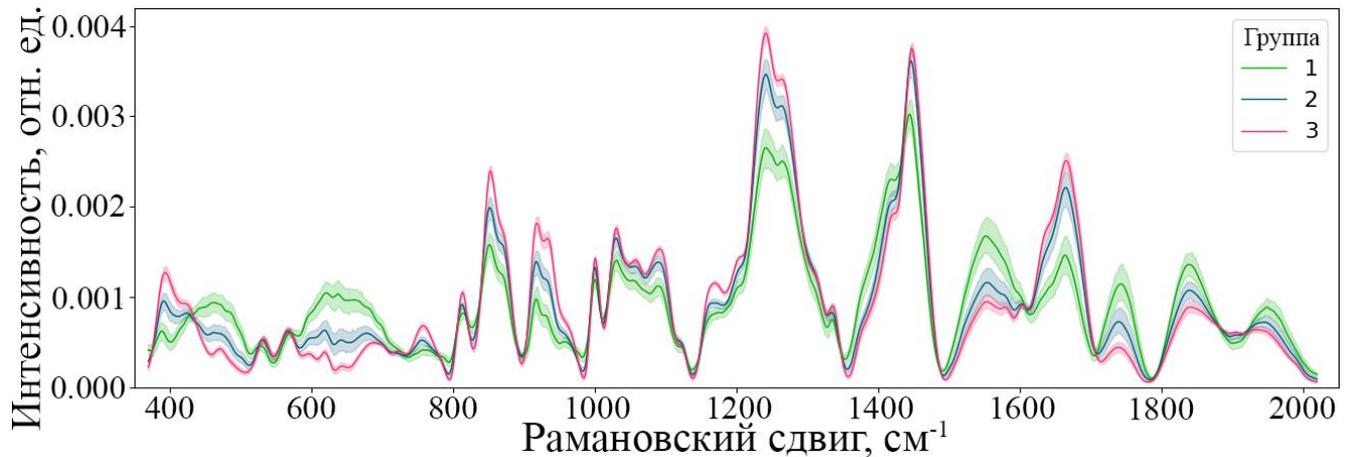


Рисунок 41 – Усредненные спектры КР исследуемых образцов: 1 – брефоматрикс из губчатой кости, 2 – брефоматрикс из кортикальной кости, 3 – биоматериалы из деминерализованных ювенильных зубов

Коэффициент корреляции Пирсона между усредненными спектрами 2 и 3 группы составляет 0,98, между 1 и 3 группой 0,84. Спектры биоматериалов из деминерализованных ювенильных зубов ближе к спектрам брефоматриксу из кортикальной кости.

В среднем относительная интенсивность основных линий возрастала от первой группы к третьей: $\sim 874 \text{ см}^{-1}$ (С-С колебания гидроксипролина), 1000 см^{-1} , 1030 см^{-1} (фенилаланин), $1200\text{--}1300 \text{ см}^{-1}$ (амид III), $\sim 1446 \text{ см}^{-1}$ (CH_2 протеин), $1633\text{--}1700 \text{ см}^{-1}$ (амид I). Эти спектральные изменения были вызваны различием органического состава испытуемых. Качественный состав линий КР спектров всех групп идентичен.

Спектры КР характеризуются отсутствием линий гидроксиапатита: ~ 432 ($\text{PO}_4^{3-} \nu_2$), $584\text{--}593$ ($\text{PO}_4^{3-} \nu_4$), $955\text{--}961$ ($\text{PO}_4^{3-} \nu_1$), ~ 1045 ($\text{PO}_4^{3-} \nu_3$) и $\sim 1075 \text{ см}^{-1}$ ($\text{CO}_3^{2-} \nu_1$).

Для декомпозиции спектров КР использован алгоритм, описанный в 3 главе. Усредненное значение метрик для декомпозированных 179 спектров КР составило: $R^2 = 99,79 \%$, $\chi^2 = 1 \cdot 10^{-6}$, $\text{aic} = -20391$, $\text{bic} = -17673$.

Усредненный спектр КР с декомпозированными линиями представлен на рисунке 42. Было выделено 185 линий КР.

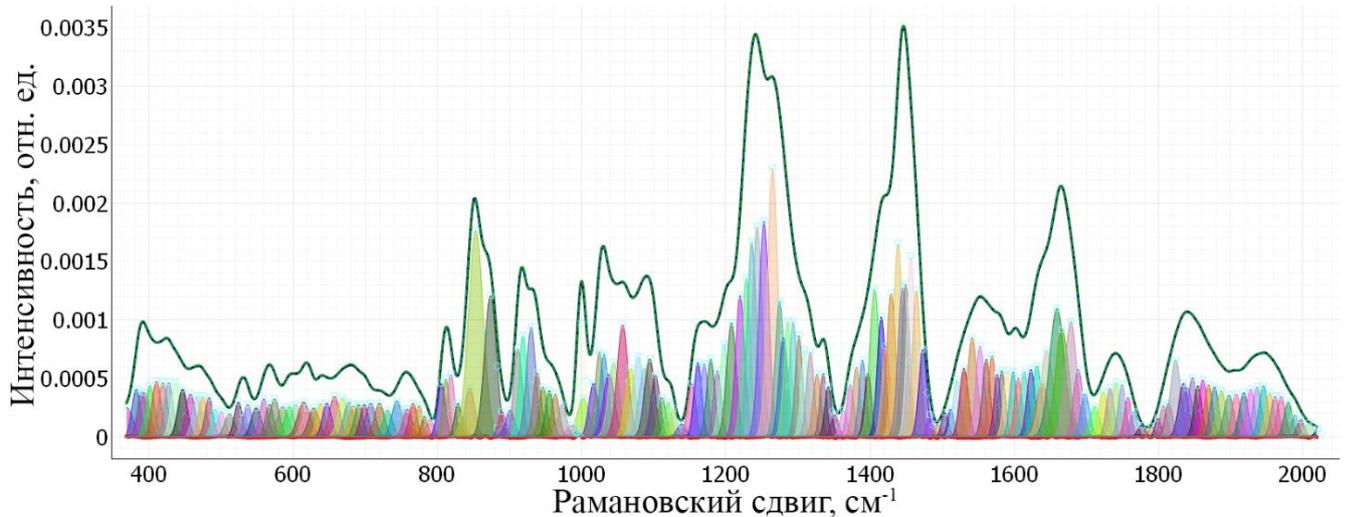


Рисунок 42 – Итоговая модель декомпозиции спектра КР в области 370 – 2020 см^{-1} . (Синяя сплошная линия – исходный спектр, зеленая – суммарный спектр элементарных линий)

4.2.3 Статистический анализ различий между группами

4.2.3.1 Проверка нормальности распределения

Перед проведением непараметрических тестов была проведена проверка нормальности распределения признаков в каждой из сравниваемых групп с использованием критерия Шапиро–Уилка. Полученные p -значения:

Для группы 1 (брефоматрикс из губчатой кости): 0,00012;

Для группы 2 (брефоматрикс из кортикальной кости): 0,00021;

Для группы 3 (биоматериалы из деминерализованных ювенильных зубов): $7 \cdot 10^{-5}$.

Обе группы демонстрируют значимые отклонения от нормального распределения (p -значение $< 0,01$), что оправдывает использование непараметрических методов сравнения, в частности – пермутационного теста.

4.2.3.2 Пермутационный тест

Результаты теста:

- Статистика теста: 0,001 (сравнение 1 и 3 группы) и 0,003 (2 и 3 группа);

- p -значение: 10^{-6} ;
- Расчетная мощность теста: 1,0;
- Необходимый минимальный размер выборки для достижения мощности 0,8: 20 спектров на группу.

Результаты свидетельствует о высокой чувствительности теста при текущем размере выборки и значительном эффекте между группами.

4.2.3.3 Распределения значений признаков

На рисунке 43 представлены значения амплитуды восьми отобранных декомпозированных линий с наибольшей дисперсией методом ANOVA (дисперсионный анализ). Видно, что для спектров 3 группы характерны более высокие значения амплитуды линии соответствующих органическим компонентам.

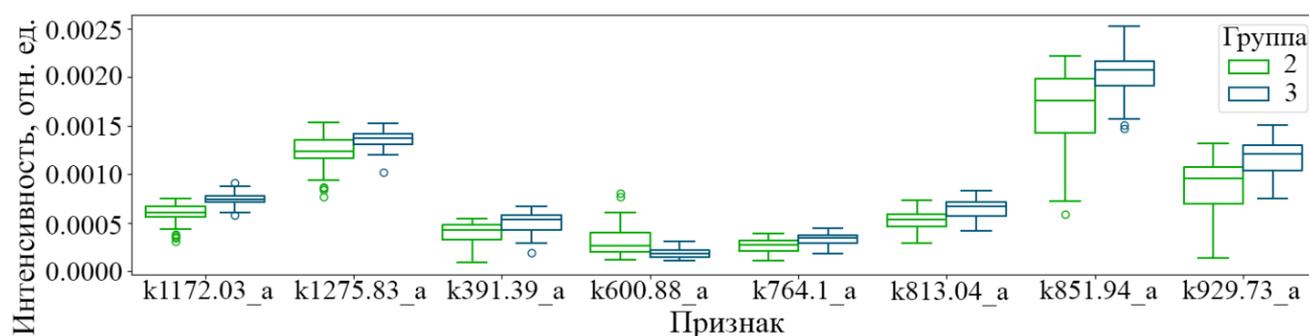


Рисунок 43 – Значения амплитуды линий ~1172, ~1276, ~391, ~601, ~764, ~813, ~852 и ~930 см^{-1} . 2 – брeфоматрикс из кортикальной кости (зеленый); 3 – биоматериалы из деминерализованных ювенильных зубов (синий)

4.2.3.4 Анализ методом главных компонент

На рисунке 44 представлен результат анализа данных МГК.

Анализ проводился с целью визуализации и интерпретации спектральных данных, полученных после декомпозиции спектров комбинационного рассеяния (КР) для сравнения биоматериалов из деминерализованных ювенильных зубов и брeфоматриксов. При переходе в пространство первых двух главных компонент распределения спектров групп 2 (брeфоматрикс из кортикальной костной ткани) и 3 (деминерализованный ювенильный дентин) продемонстрировали значительное перекрытие, что указывает на высокую степень схожести их качественного состава.

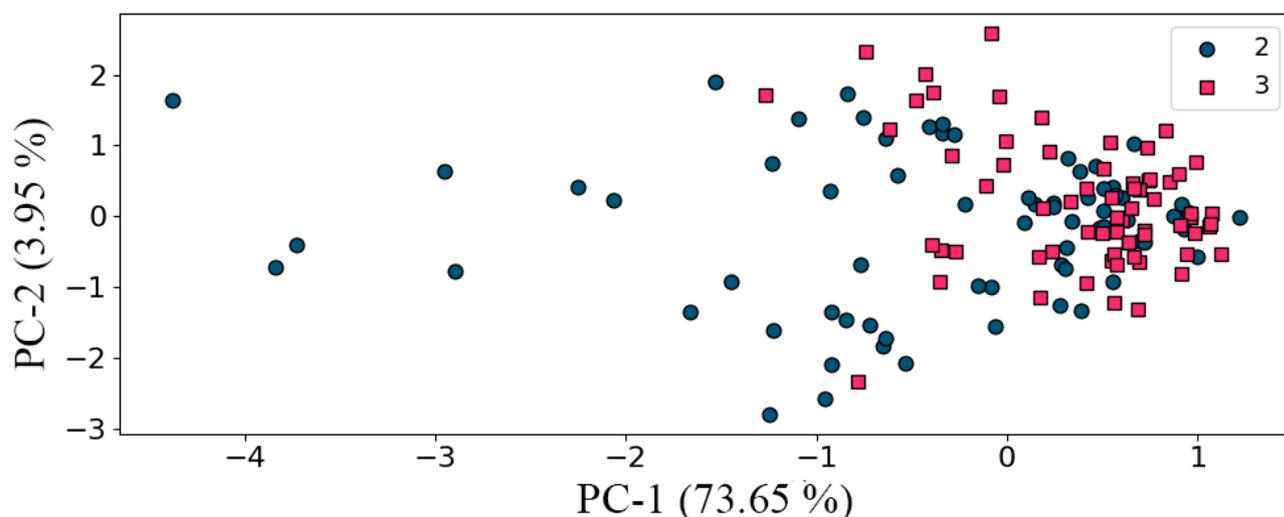


Рисунок 44 – График решений модели для данных в пониженной размерности. 2 – брeфоматрикс, 3 – деминерализованный ювенильный дентин

Корреляционный анализ подтверждает данное наблюдение: коэффициент корреляции Пирсона между усреднёнными спектрами групп 2 и 3 составил 0,98, в то время как между группами 1 и 3 – 0,84. Эти результаты свидетельствуют о том, что спектральные характеристики, определяемые по относительной интенсивности декомпозированных линий КР, существенно зависят от качественного состава исследуемых материалов, и что биоматериалы из ювенильных зубов по спектральному содержанию ближе к брeфоматриксу из кортикальной кости.

4.3 Выводы по главе.

1. Продемонстрировано, что спектроскопия комбинационного рассеяния в сочетании с разработанными алгоритмами обработки спектров позволяет детектировать изменения состава дентина при деминерализации. В эксперименте было проанализировано 344 спектра (194 – для минерализованного дентина и 150 – для деминерализованного). Изначально выделено 167 спектральных линий, из которых с помощью ANOVA и рекурсивного удаления признаков отобраны 83, а затем 5 ключевых, соответствующих амплитудам линий при 956, 1244, 1446, 1664 и ~ 870 см^{-1} .

Статистический анализ подтвердил значимость различий: для группы минерализованного дентина получено p -значение (p -value, отражающее

вероятность случайного совпадения при верной нулевой гипотезе) = 0,0005, а для деминерализованного – $9,37 \cdot 10^{-6}$. Пермутационный тест показал статистику 0,0101 с p -значением, приближаемым к 0, а расчетная мощность теста составила 1,0 (минимальное число спектров для достижения мощности 0,8 – 5 на группу).

Построенная модель логистической регрессии, использующая выбранные 5 признаков, продемонстрировала высокие значения метрик: точность (accuracy), ROC AUC, точность (precision), полнота (recall) и F1-мера – все на уровне 1,0 (при LogLoss около 0,015112 на обучающей выборке и 0,013082 на тестовой). SHAP-анализ дополнительно выявил, что признак с амплитудой 956 см^{-1} оказывает инвертированное влияние (отрицательный коэффициент -1.97) по сравнению с остальными положительными коэффициентами, что позволяет интерпретировать промежуточные значения деминерализации с высокой достоверностью. Эти результаты демонстрируют потенциал способа для точного контроля качества дентина в клинических приложениях.

2. В рамках сравнительного анализа спектральных характеристик использовались образцы трех групп: 36 образцов брэфоматрикса из губчатой кости, 36 – из кортикальной кости, и 47 – из деминерализованного ювенильного дентина. Усредненные спектры, обработанные по разработанному способу, показали, что биоматериалы из ювенильного дентина по спектральным характеристикам очень близки к брэфоматриксу из кортикальной кости: коэффициент корреляции Пирсона составил 0,98 (для брэфоматрикса из кортикальной кости и ювенильного дентина) против 0,84 (между брэфоматриксом из губчатой кости и дентина).

Декомпозиция спектров позволила выделить 185 линий с высокими показателями: коэффициент детерминации $R^2 = 0,99$, значение $\chi^2 = 1 \cdot 10^{-6}$, AIC = - 20391, BIC = -17673. Статистический анализ различий между группами методом пермутационного теста выявил тестовые статистики 0,001 (сравнение брэфоматрикса из губчатой кости и ювенильного дентина) и 0,003 (брэфоматрикс из кортикальной кости и ювенильного дентина) с p -значением порядка 10^{-6} , что подтверждает высокую значимость различий.

Кроме того, анализ спектральных линий (в диапазонах ~ 391 , ~ 601 , ~ 764 , ~ 813 , ~ 852 , ~ 873 , ~ 930 , ~ 1000 , ~ 1030 см^{-1} , $1200\text{--}1300$ см^{-1} , ~ 1446 и $1633\text{--}1700$ см^{-1}) показал, что несмотря на различия в интенсивности отдельных линий, качественный состав всех групп идентичен. При этом биоматериалы из ювенильных зубов характеризуются более высокими значениями линий, ассоциированных с органическим компонентом (например, линии около 873 см^{-1} , отражающие C–C колебания гидроксипролина), что обуславливает их оптимальное применение в трансплантационной практике.

5 ЭКСПРЕСС-ДИАГНОСТИКА ПАРОДОНТИТА И ОЦЕНКА ЕГО ЛЕЧЕНИЯ С ПОМОЩЬЮ РАЗРАБОТАННОГО СПОСОБА

Пятая глава посвящена изучению возможности диагностики пародонтита и оценки его лечения с помощью спектроскопии комбинационного рассеяния и разработанного способа обработки спектров КР.

Цель данной главы – оценить возможность диагностики признаков пародонтита с помощью метода спектроскопии КР с применением разработанного способа и методом машинного обучения.

5.1 Оптическая диагностика изменений эмали зубов при пародонтите

В разделе 5.1 представлены результаты экспериментальных исследований эмали зубов при заболевании пародонтит с помощью способа спектроскопии КР. Изучены спектральные изменения тканей зубов при пародонтите. Разработаны критерии для неинвазивной диагностики пародонтита по изменению спектральных свойств эмали зубов.

Цель раздела – оценка возможности использования предложенного способа анализа спектров КР с применением методов машинного обучения для диагностики признаков пародонтита в эмали.

5.1.1 Пробоподготовка и алгоритмы обработки спектров КР

Материалами исследований являлись 42 зуба (моляры, премоляры и клыки) пациентов обоего пола, возрастной группы от 35-70 лет, удаленные в связи с хроническими пародонтитами и по ортодонтическим показаниям (контрольная группа). Диагноз пародонтит ставился клинически и при анализе компьютерной конусно-лучевой томографии (КТ) (код заболевания по МКБ-10 (1997 г.) - K05.3). Для основной группы исследования отбирались зубы, удаленные по поводу хронического пародонтита тяжелой степени, с глубиной пародонтальных карманов от 6 мм и более и патологической подвижностью III-IV степени. На компьютерных томограммах убыль костной ткани вокруг корней удаленных зубов составила более $\frac{1}{2}$ длины корня.

В каждой исследуемой области были сняты по 3 спектра в 3-5 различных точках поверхности каждой тканей зуба. Все образцы были поделены на 2 основные группы: I – группа с диагнозом пародонтит, II – контрольная группа.

Также были проведены исследования *in vivo* эмали 22 зубов добровольца (моляров, премоляров и клыков) в рамках одного пациента на ранней стадии развития пародонтита. В данных исследованиях принимал участие доброволец, женского пола. В полости рта пациента зуб с установленным диагнозом – локальный пародонтит (код заболевания по МКБ-10 (1997 г.) - K05.3) (рисунок 45).



Рисунок 45 – Процесс регистрации спектров *in vivo*

Итоговое количество спектров КР 1 группы – 63, 2 группы – 82.

Исследования проводились с помощью метода спектроскопии комбинационного рассеяния, который был реализован с помощью экспериментального стенда.

Данный экспериментальный стенд включал в себя полупроводниковый лазер (LML-785.0RB-04), оптический модуль комбинационного рассеяния (PBL 785), спектрограф (Sharmrock SR-303i) с интегрированной цифровой камерой (ANDOR DV-420A-OE), охлаждаемой до -60°C , и компьютер.

Использование данного спектрографа обеспечивало разрешение 0,15 нм по длине волны при низком уровне собственных шумов. В данной работе анализ КР спектров проводился в диапазоне $320-2200\text{ см}^{-1}$. Мощность излучения лазера 400 мВт. Регистрация спектров КР проводилась с помощью оптического зонда, который располагался над объектом на расстоянии 7 мм.

Полученные спектры комбинационного рассеяния далее обрабатывались алгоритмами в соответствии с блок-схемой на рисунке 46 до этапа получения таблицы с амплитудами декомпозированных линий КР.

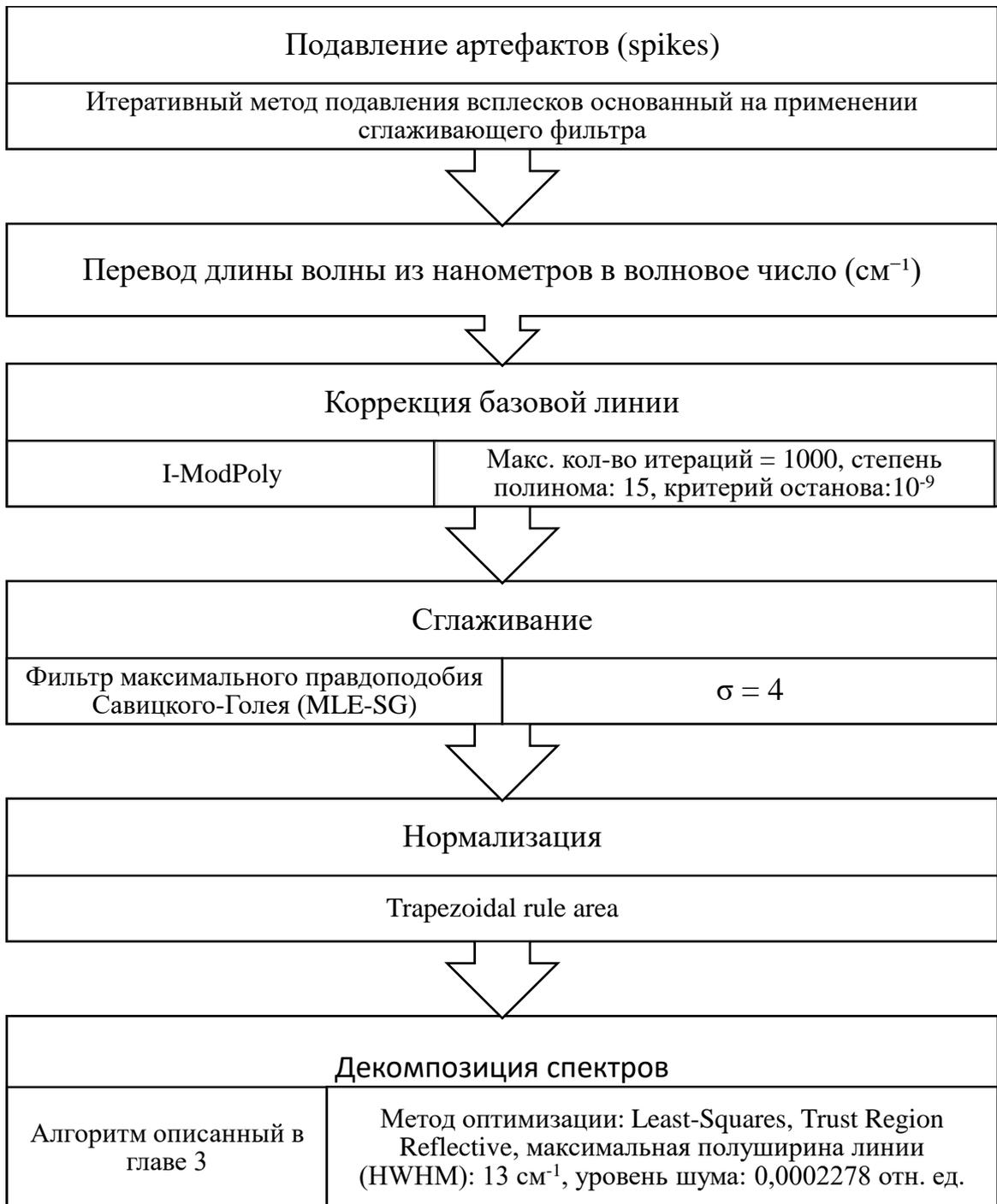


Рисунок 46 – Блок схема математической обработки спектров КР

5.1.2 Анализ усредненных спектров КР и результаты декомпозиции

На рисунке 47 и 48 приведены усредненные спектры эмали зубов при пародонтите (2) и контрольная группа (1). Анализ спектров здоровых тканей зубов

и тканей зубов при пародонтите показывает, что основные спектральные различия тканей зубов при пародонтите проявляются в области спектра 630 – 690 (деформационные моды PO_4^{3-} ν_4), ~850 (C–C колебания пролина и гидроксипролина), 1230 – 1280 (амид III (сгибание C–N, N–H), ~1319 (комбинационные или боковые колебания органических фракций (амид III) [134] или Guanine (B, Z-marker) [135]), 1420 – 1460 (CH_2 деформационные колебания протеина) [130], ~1552 (амид II колебания (C–N растяжение, N–H сгибание) органических компонентов), 1610 – 1670 (амид I) и ~1745 cm^{-1} (C=O растяжение органических фракций или карбонатных групп) [136].

Изменений амплитуды интенсивности линий гидроксиапатита не наблюдается: ~957 (P–O симметричное валентное PO_4^{3-} (ν_1)), ~430 (PO_4^{3-} (ν_2) [129]), ~586 (PO_4^{3-} (ν_4) [129]), ~1043 (PO_4^{3-} (ν_3) [129]), ~1068 (C–O плоскостное валентное, CO_3^{2-} (ν_1) замещение B-типа [137]) cm^{-1} .

Также статистически значимых отличий в относительной интенсивности других линий КР не наблюдается. Коэффициент корреляции Пирсона между усредненными спектрами КР двух групп = 0,990.

Данные спектральные проявления, по-видимому, связаны с биохимическими процессами, происходящими во время появления поверхностных образований зубов при пародонтите (зубные камни, бляшки и т.д.), что также отражается на всех тканях зубов. Этиология образования зубного камня связана с механизмами минерализации поверхностных образований на зубе, которые состоят из углеводов и протеинов (по 30 %), а также в них содержится около 15 % липидов. Остальные компоненты представлены внеклеточными продуктами жизнедеятельности бактерий бляшки, остатками их цитоплазмы и клеточной мембраны (экстрацеллюлярными полисахаридами) [138].

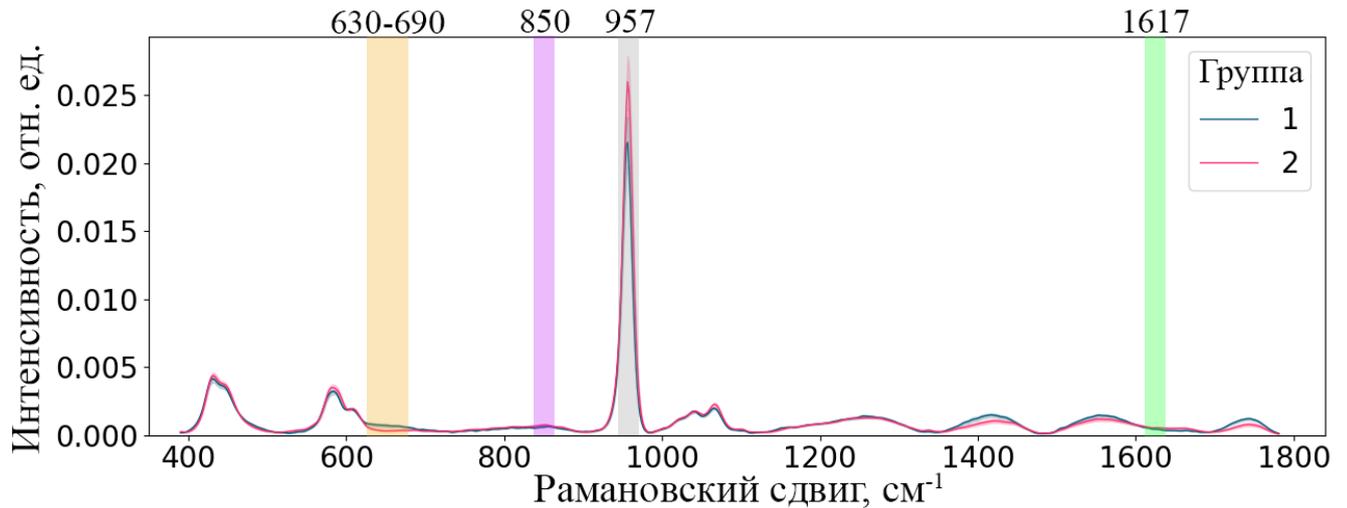


Рисунок 47 – Усредненные спектры КР эмали зубов: 1 – контрольная группа, 2 – с диагнозом пародонтит

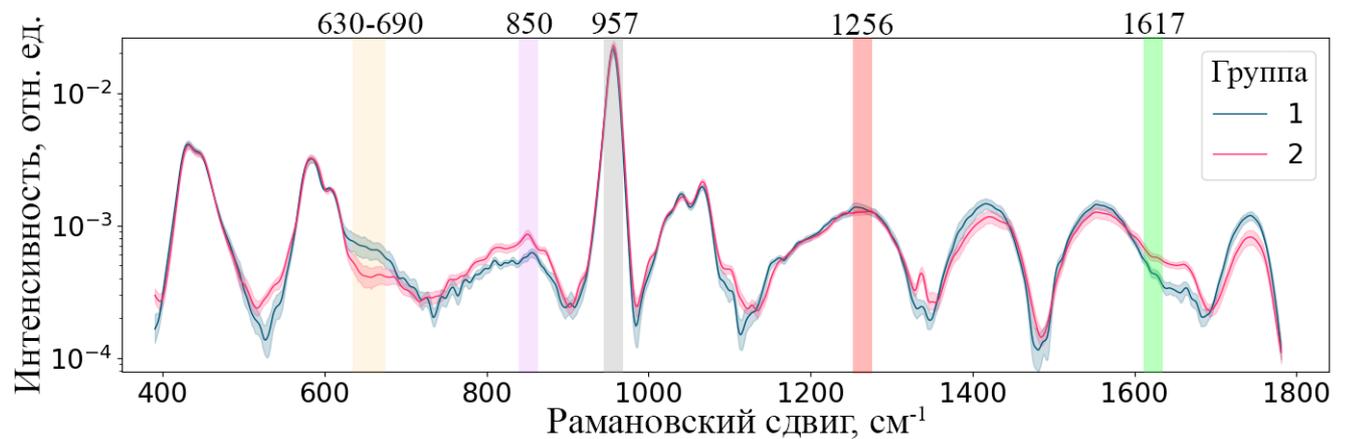


Рисунок 48 – Усредненные спектры КР эмали зубов в логарифмическом масштабе интенсивности: 1 – контрольная группа, 2 – с диагнозом пародонтит

Главными неорганическими веществами зубного камня являются кальций (39 %), фосфор (19 %), карбонаты (1,9 %). Из микроэлементов присутствуют Na^+ , Zn^{2+} , Cu^{2+} , Mn^{2+} , Ag^+ , Al^{3+} , Fe^{2+} , F^- . [31].

Для декомпозиции спектров КР использован алгоритм, описанный в 3 главе. Усредненное значение метрик для декомпозированных 145 спектров КР составило: $R^2 = 99,92 \%$, $\chi^2 = 2 \cdot 10^{-6}$, $\text{aic} = -15685$, $\text{bic} = -13681$. Полученные показатели свидетельствуют о том, что модель, полученная МНК, объясняет данные на приемлемом уровне. Можно отметить следующие аспекты:

- Хи-квадрат ($\text{Chi-square } 2 \times 10^{-6}$) и сокращённый хи-квадрат ($\text{reduced chi-square} \approx 10^{-5}$): Низкие значения говорят о том, что сумма квадратов отклонений

между экспериментальными данными и моделью крайне мала, а среднее отклонение (с учётом числа степеней свободы) практически не наблюдается. Это указывает на очень хорошее соответствие модели данным.

- Коэффициент детерминации ($R^2 \approx 0,9992$): практически единичное значение коэффициента детерминации означает, что модель объясняет более чем 99,9 % варибельности экспериментальных данных.

- AIC ≈ -15685 , BIC ≈ -13681 : Отрицательные и достаточно низкие значения информационных критериев указывают на то, что выбранная модель лучше описывает данные по сравнению с потенциальными альтернативными моделями, с учётом штрафа за число параметров.

В совокупности эти результаты говорят о том, что декомпозиция спектров КР на 142 линий гаусса посредством МНК выполнена с высоким качеством. Однако стоит учитывать, что крайне низкое значение reduced chi-square (близкое к нулю) может быть связано как с чрезвычайно хорошим соответствием, так и с возможной переоценкой погрешностей измерений. Если допущения о распределении ошибок и оценке неопределённостей корректны, можно утверждать, что результаты декомпозиции являются очень хорошими.

5.1.3 Отбор признаков

На первом этапе отбора признаков амплитуды декомпозированных линий были проанализированы методом ANOVA (Дисперсионный анализ с использованием класса SelectPercentile библиотеки sklearn с параметром percentile = 30) и из 284 было отобрано 85 признаков.

На втором этапе отбора признаков была применена встроенная функция select_features градиентного бустинга на решающих деревьях CatBoostClassifier, ориентированная на итеративное исключение признаков с наименьшим вкладом в предсказательную способность модели. Метод реализует разновидность рекурсивного исключения признаков (RFE, Recursive Feature Elimination) с использованием собственной метрики качества модели.

Для сохранения сбалансированности классов в условиях дисбаланса была применена параметрическая настройка `scale_pos_weight`, компенсирующая различия в численности классов. В качестве метрики оценки качества использовалась площадь под ROC-кривой (AUC), а обучение производилось на GPU для повышения вычислительной эффективности.

Алгоритм последовательно исключал признаки с наименьшим вкладом, производя оценку качества модели на отложенной валидационной выборке (20 % от исходных данных, стратифицированной по целевой переменной). В результате данного этапа был сформирован поднабор из 20 признаков, обладающих наибольшей значимостью с точки зрения повышения обобщающей способности модели.

На заключительном этапе отбора признаков был проведён корреляционный анализ с целью исключения избыточных (высококоррелированных) признаков, способных вносить мультиколлинеарность в модель (рисунок 49).

Дополнительно на данном этапе были внесены экспертные корректировки на основе предварительных результатов моделирования. В результате итоговый набор был сокращён до 4 информативных и взаимно слабо коррелированных признаков, представляющих наибольший интерес как с точки зрения статистической значимости, так и с точки зрения интерпретируемости модели., соответствующих амплитуде декомпозированных линий на ~ 649 , ~ 850 и ~ 1618 см^{-1} и положению максимума линии 957 см^{-1} .

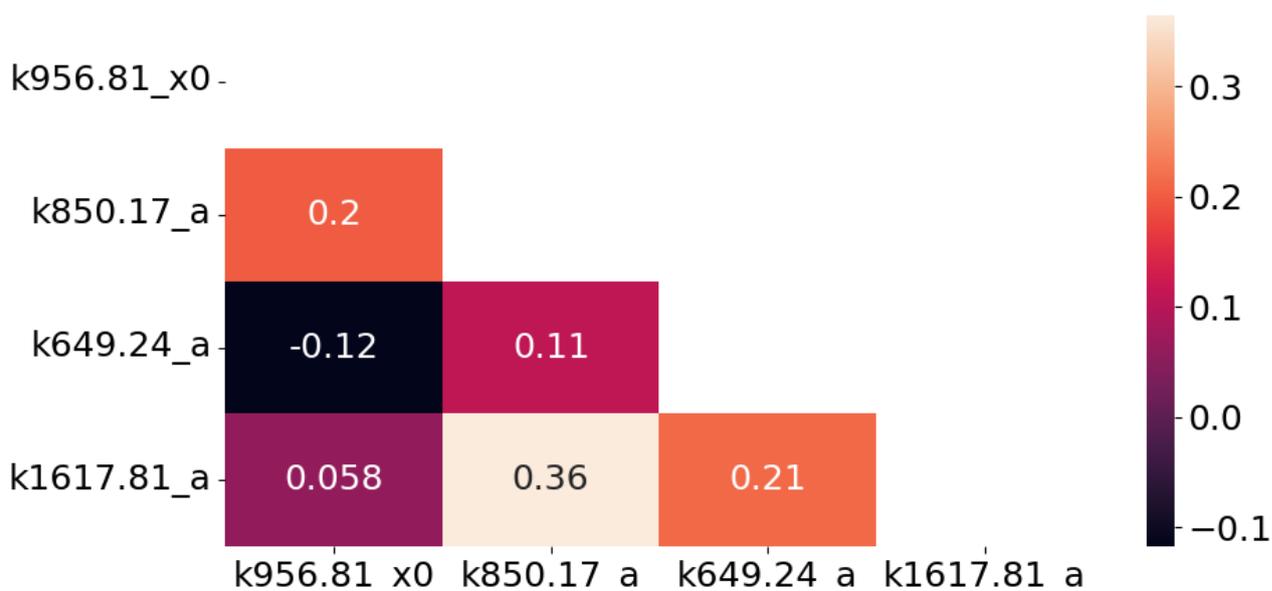


Рисунок 49 – Корреляционная матрица отобранных признаков

5.1.4 Статистический анализ различий между группами

5.1.4.1 Проверка нормальности распределения

Перед проведением непараметрических тестов была проведена проверка нормальности распределения признаков в каждой из сравниваемых групп с использованием критерия Шапиро–Уилка. Полученные значения p -значений:

Для группы 1 (контрольная группа): $2 \cdot 10^{-3}$;

Для группы 2 (эмаль с диагнозом пародонтит): $1 \cdot 10^{-13}$.

Обе группы демонстрируют значимые отклонения от нормального распределения (p -значение < 0.01), что оправдывает использование непараметрических методов сравнения, в частности – пермутационного теста и критерия Манна–Уитни.

5.1.4.2 Пермутационный тест

Результаты теста:

- Статистика теста: 0,41;
- p -значение: 10^{-5} ;
- Расчетная мощность теста: 1,0;
- Необходимый минимальный размер выборки для достижения мощности 0,8: 7 спектров на группу.

Результаты свидетельствует о высокой чувствительности теста при текущем размере выборки и значительном эффекте между группами.

5.1.4.3 Критерий Манна–Уитни (U-тест)

U-тест был применён к каждому признаку по отдельности. Для каждой переменной рассчитывались: статистика U и соответствующее p -значение.

U-статистика (U-value) – это величина, рассчитываемая в рамках теста Манна–Уитни, которая отражает степень различия между выборками. Чем ниже значение U при фиксированном размере выборок, тем больше различие между группами. Однако интерпретация результата осуществляется на основе соответствующего p -значения.

В таблице 8 представлены значения U-статистики и соответствующих уровней значимости (p -значение) для каждого из 4 признаков.

Таблица 8. Результаты U-теста

Признак	U-статистика	p -значение
k649.24_a	33439,5	10^{-5}
k850.17_a	10359,0	10^{-5}
k956.81_x0	8833,0	10^{-5}
k1617.81_a	13737,5	10^{-5}

Во всех случаях наблюдаются p -значения меньше 0,01, что свидетельствует о наличии статистически значимых различий между группами по каждому из исследуемых признаков.

Дополнительно результаты U-теста были подтверждены с использованием пермутационного теста, что позволило исключить влияние случайных факторов и повысить доверие к полученным результатам. Совпадение результатов обоих методов и наличие устойчиво низких p -значений указывает на высокую статистическую мощность анализа и подтверждает, что объём выборки является достаточным для выявления существующих эффектов.

5.1.4.4 Распределения значений признаков

Рисунок 50 демонстрирует распределения амплитуд наиболее информативных линий КР, отобранных на основании результатов статистического анализа. Графики построены отдельно для контрольной группы (отмечена синим цветом) и группы пациентов с диагнозом пародонтит (отмечена красным цветом). Наблюдается отчётливая разница в распределениях, что указывает на возможные структурные и/или биохимические изменения в составе эмали зубов.

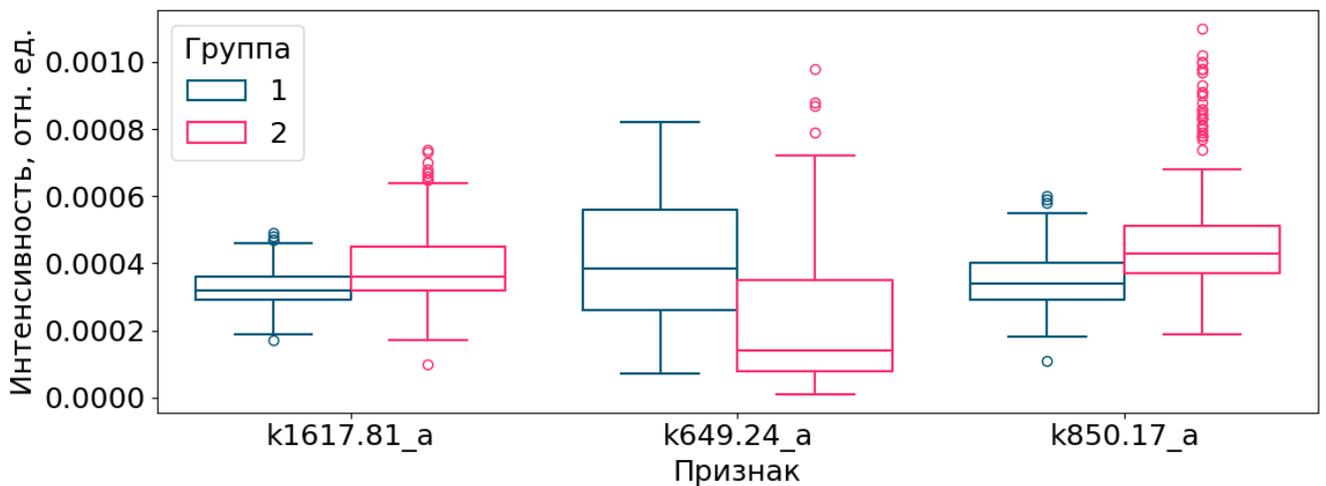


Рисунок 50 – Распределения амплитуды наиболее информативных линии КР: 1 – контрольная группа, 2 – с диагнозом пародонтит

Рисунок 51 иллюстрирует распределение положений линии КР $\sim 957 \text{ см}^{-1}$. Данный параметр отражает положение линий КР, чувствительных к локальному химическому окружению и кристаллическому напряжению. У образцов из группы с диагнозом пародонтит значения положения линии смещены в сторону более высоких волновых чисел (в среднем $956,9 \text{ см}^{-1}$ против $956,5 \text{ см}^{-1}$ в контрольной группе), что может быть связано с изменениями в кристаллической решётке гидроксиапатита под воздействием патологического процесса. Различия между группами статистически значимы (СКО: 0,32 и 0,36 соответственно), что подчёркивает диагностическую ценность данного признака.

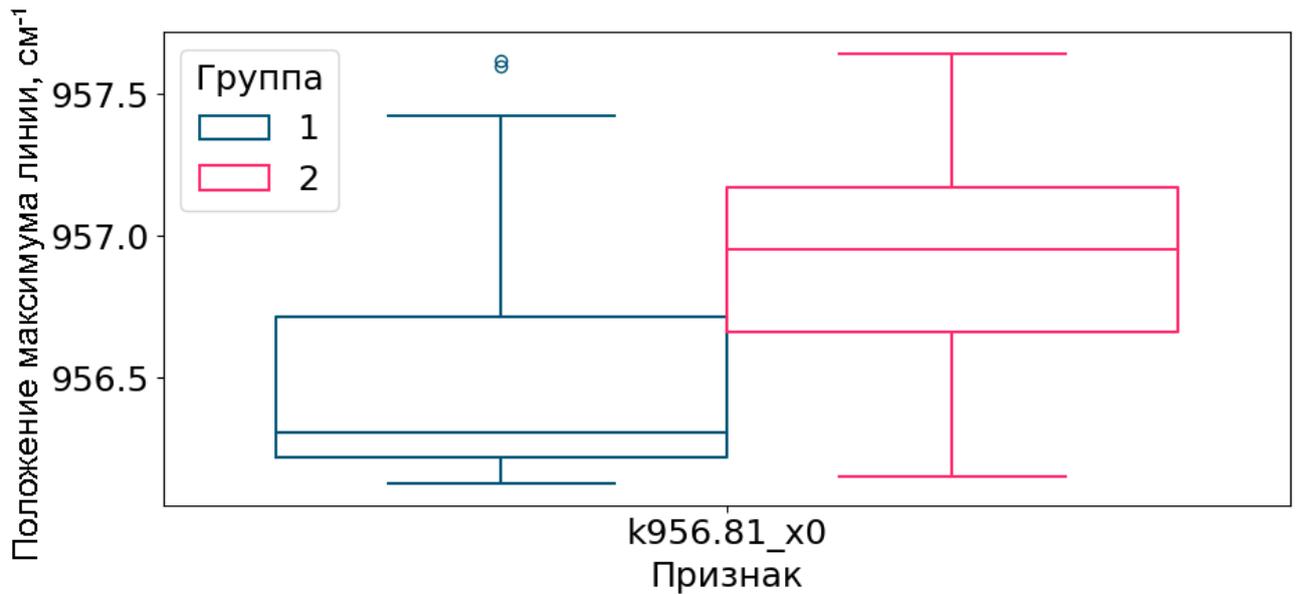


Рисунок 51 – Значения положения линии КР ~ 957 cm^{-1} : 1 – контрольная группа, 2 – с диагнозом пародонтит

5.1.5 Процесс создания модели машинного обучения

В рамках данного исследования в качестве базовой модели была выбрана логистическая регрессия. Выбор обоснован ранее в разделе 4.1.5.1.

5.1.5.1 Подготовка выборки для обучения модели

Для повышения объема обучающей выборки, увеличения обобщающей способности и обеспечения устойчивости моделей машинного обучения к ограниченности данных была реализована процедура синтетической генерации данных с применением модели Conditional Tabular GAN (CTGAN - генеративно-сопоставительная сеть для работы с табличными данными). Генерация проводилась отдельно для каждого класса с учетом специфики признаков. Для каждого подмножества выполнялось масштабирование признаков в диапазон $[0, 1]$.

Для каждого класса отдельно обучалась модель CTGAN (10000 эпох, размер батча = 64). Модель генерировала синтетический набор объемом, в два раза превышающим размер оригинального подмножества. Полученные данные масштабировались обратно к исходному диапазону.

Для каждого класса и каждого признака проводилось визуальное сравнение распределений между оригинальными и синтетическими данными. Это позволило выявить отклонения и визуально оценить адекватность сгенерированной выборки.

Для построения модели использовались данные, представляющие таблицу, в которой каждый спектр характеризуется пятью признаками, отражающими амплитуду декомпозированных линий КР на ~ 649 , ~ 850 и ~ 1618 см^{-1} и положению максимума линии 957 см^{-1} .

Спектры были распределены по двум группам:

- Группа 1 (контрольная группа): 222 спектра,
- Группа 2 (эмаль с диагнозом пародонтит): 193 спектра.

Несмотря на некоторое числовое различие между группами, дисбаланс классов является незначительным, что обеспечивает корректное представление распределения классов как в обучающей, так и в тестовой выборках.

Выборка разделялась на обучающую и тестовую случайным образом в отношении 75:25 с использованием стратификации по целевому признаку.

Перед обучением модели данные подвергаются дополнительной обработке, направленной на улучшение качества обучения. В рамках конвейера обработки используется `QuantileTransformer`.

5.1.5.2 Обучение модели логистической регрессии

Для построения базовой модели бинарной классификации использовалась логистическая регрессия, реализованная в рамках конвейера (`Pipeline`), включающего этап предварительной обработки признаков с использованием `QuantileTransformer` и калибровки модели с помощью сигмоиды. При выборе модели логистической регрессии также учитывалось, что применение других методов: SVM с нелинейным ядром, решающее дерево, случайный лес, `CatBoost` существенного прироста качества модели не помогло достигнуть.

Обучение модели сопровождалось оптимизацией гиперпараметров алгоритмом байесовской оптимизации с помощью библиотеки `Optuna`. Целевой функцией (`objective`) являлось среднее значение площади под ROC-кривой (ROC AUC), рассчитанное на основе стратифицированной кросс-валидации с 3 фолдами и Оценка Брайера (`Brier-Score`). Такой подход обеспечивает устойчивую оценку

обобщающей способности модели, калибровки и снижает риск переобучения, особенно при ограниченном объёме данных.

Целевая функция использует выбор гиперпараметров:

- типа регуляризации (penalty),
- метода оптимизации (solver),
- коэффициента регуляризации (C),
- параметра l1_ratio для ElasticNet (при необходимости).

Параметры выбирались из допустимых комбинаций, предусмотренных библиотекой sklearn, с учётом совместимости между типом регуляризации и решателем. Также в модели используется параметр class_weight='balanced', позволяющий автоматически компенсировать незначительное смещение классов, тем самым обеспечивая равное внимание обоим классам в процессе обучения.

Найденные оптимальные значения приведены в таблице 9.

Таблица 9. Гиперпараметры модели и их значения

Параметр	Значение	Обоснование
penalty	'l2'	L2 предпочтительна, т.к. она устойчива к мультиколлинеарности.
C	0,081	Значение 0.081 говорит о достаточно сильной регуляризации, что логично для небольших выборок
Solver_l2	liblinear	Работает хорошо при разреженных или немногочисленных признаках

5.1.6 Оценка качества и интерпретация полученного классификатора

Таблица 10 содержит основные метрики качества модели, полученной на выборках, состоящих из 311 спектров для обучения и 104 спектров для валидации. Значения точности (accuracy) составляют 0,90 для обучающей выборки и 0,88 для тестовой, что свидетельствует о высоком проценте правильных предсказаний при отсутствии снижения эффективности на валидационных данных. Показатель ROC AUC, равный 0,95 и 0,94 соответственно, подтверждает высокую чувствительность

и специфичность модели, что является важным для задач бинарной классификации при диагностике пародонтита.

Анализ значений precision, recall и F1-меры также позволяет отметить, что модель обучена стабильно, не переобучена, отлично обобщается на новые данные и даёт уверенные предсказания.

Таблица 10. Значения метрик модели логистической регрессии, доверительные интервалы рассчитаны для уровня вероятности не менее 95 %

Метрика	Обучающая выборка (311 спектров)	Тестовая выборка (104 спектра)
Точность (accuracy)	0,90 (0,86-0,93)	0,88 (0,82-0,94)
ROC AUC	0,95 (0,93-0,97)	0,94 (0,89-0,98)
Точность (precision)	0,93 (0,84-0,98)	0,90 (0,85-0,94)
Чувствительность (recall)	0,91 (0,86-0,95)	0,86 (0,76-0,94)
Специфичность	0,91 (0,86-0,95)	0,86 (0,76-0,94)
Оценка F1 (F-мера)	0,90 (0,87-0,94)	0,89 (0,82-0,94)
Логарифмическая потеря (Logloss)	0,31 (0,26-0,35)	0,33 (0,26-0,42)

Распределение скоров модели (score) (рисунок 52) демонстрирует определённое перекрытие, однако наблюдаемое смещение средних значений указывает на способность модели учитывать различия между классами. Этот результат подтверждает, что, несмотря на перекрытие, входные признаки обладают диагностической информацией, позволяющей классификатору выделять патологические состояния.

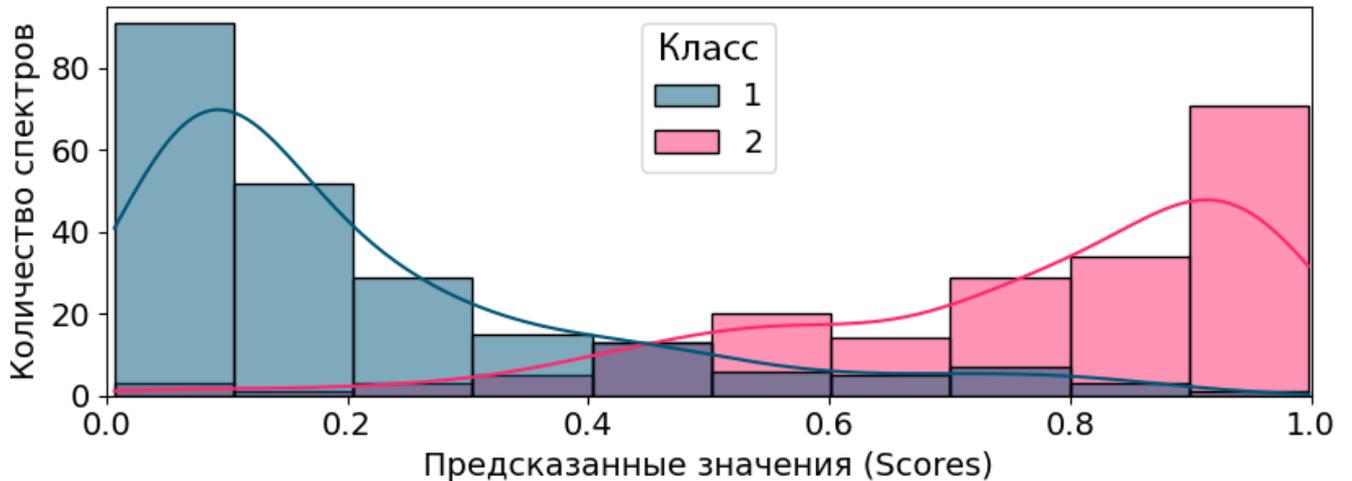


Рисунок 52 – Выходные значения модели (score). 1 – контрольная группа, 2 – с диагнозом пародонтит

На проекции с использованием МГК (рисунок 53) наблюдается пересечение классов. Видно, что первая главная компонента (PC-1) объясняет 99,96 % дисперсии, что указывает на практическую одномерность данных.

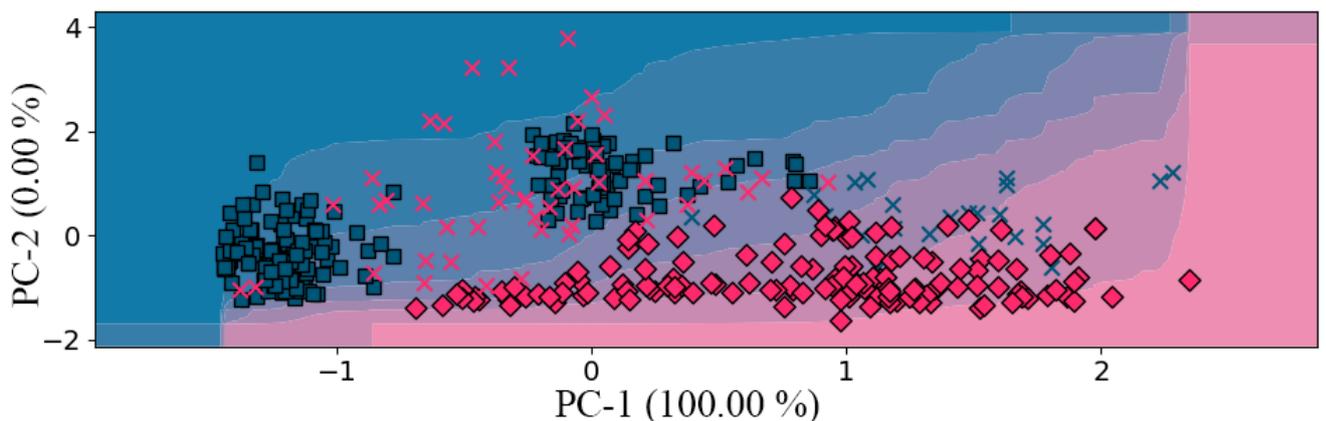


Рисунок 53 – Визуализация границ принятия решения модели логистической регрессии в пространстве главных компонент. Синий – контрольная группа, красный – с диагнозом пародонтит

Наблюдается, что логистическая регрессия формирует линейно разделяющее пространство, соответствующее характеру признаков. Наличие перекрытия между классами в области около нуля по PC-1 указывает на частичную неразделимость классов и возможное присутствие шумовых наблюдений.

Матрица ошибок (рисунок 54) подтверждает, что классификатор обладает высокой способностью к различению классов, демонстрируя устойчивость и сбалансированную производительность. Это подтверждает ранее рассчитанные

метрики (accuracy, precision, recall, F1), и свидетельствует о надёжности модели в практическом применении.



Рисунок 54 – Матрица решений

Кривая ошибок первого и второго рода (ROC AUC curve) показана на рисунке 55, площадь под кривой (AUC) составляет 0,971, что говорит о высокой классифицирующей способности модели.

Анализ кривой точности – полноты (Precision-Recall curve) (рисунок 55) показал, что логистическая регрессия достигает высокого значения средней точности (Average Precision) ($AP = 0,92$), что свидетельствует о низком количестве ошибок классификации.

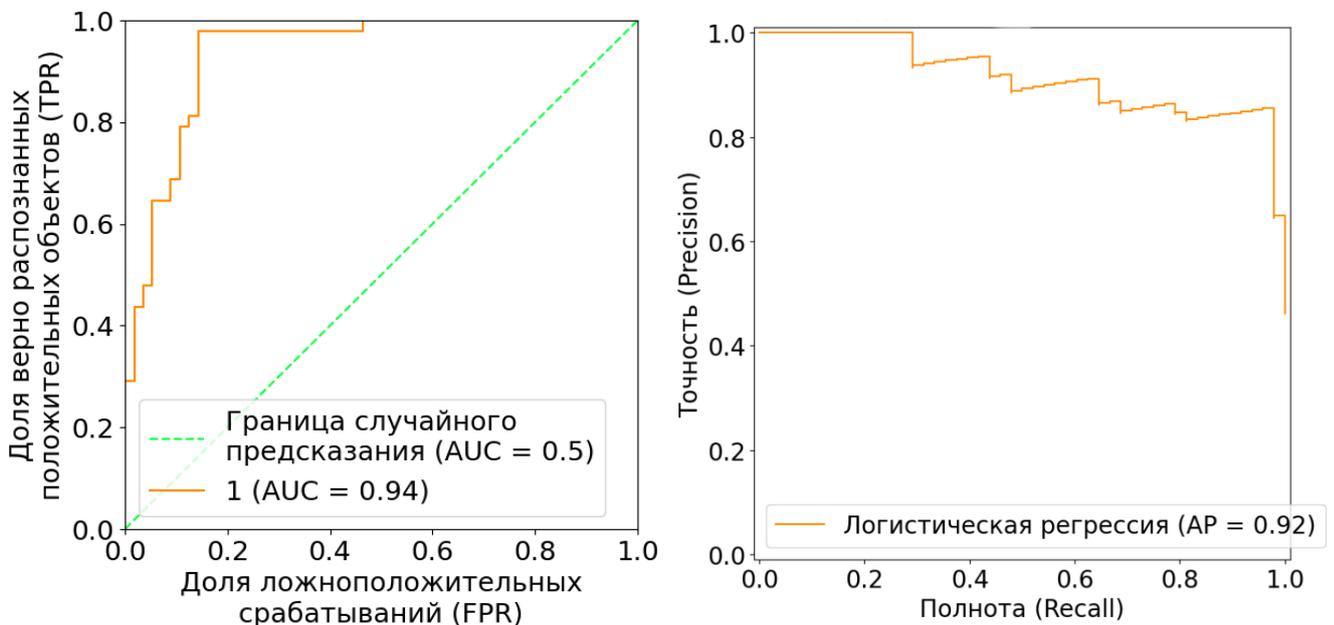


Рисунок 55 – Кривая ошибок первого и второго рода (ROC AUC curve) (слева) и кривая точности – полноты (Precision – Recall curve) (справа)

Для интерпретации вклада отдельных признаков в предсказания модели логистической регрессии был построен график значений SHAP (рисунок 56), отражающий влияние каждого признака на выход модели.

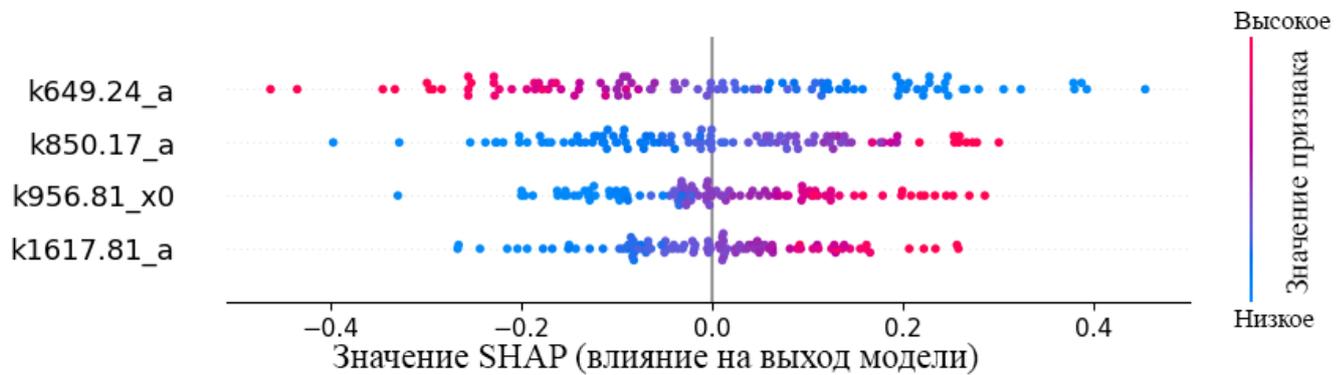


Рисунок 56 – график SHAP-значений для логистической регрессии. По оси Y отложены линии КР, по оси X – величина их вклада в предсказание модели (SHAP).

Красный цвет соответствует высоким значениям признака, синий – низким. Положительные SHAP-значения увеличивают вероятность прогноза как целевого класса (пародонтит), отрицательные – уменьшают.

Суммарно, видно, что признаками пародонтита являются высокие значения признаков, соответствующих положению максимума линии 957 см^{-1} и амплитуды линий 850 и 1618 см^{-1} . И соответственно более низкие значения линии 649 см^{-1} .

На тепловой карте (рисунок 57) наблюдения (спектры по оси X) упорядочены по значению функции модели $f(x)$, то есть по выходу модели. Это видно из плавного графика над тепловой картой, который отражает последовательность предсказаний – от высоких значений к низким (или наоборот), позволяя сопоставить изменение SHAP-вкладов с изменением предсказания.

Логика модели устойчива. Наблюдается систематическое поведение SHAP-значений вдоль оси X – например, на высоких значениях $f(x)$ преобладает красный цвет, а на низких – синий. Это означает, что влияние признаков согласуется с предсказанием модели, и она ведёт себя логично.

Поскольку модель уверенно усиливает или подавляет вероятность предсказания с помощью разных признаков (в зависимости от спектра), можно

заклучить, что она обладает дискриминационной способностью и выявляет паттерны, характерные для целевого класса.

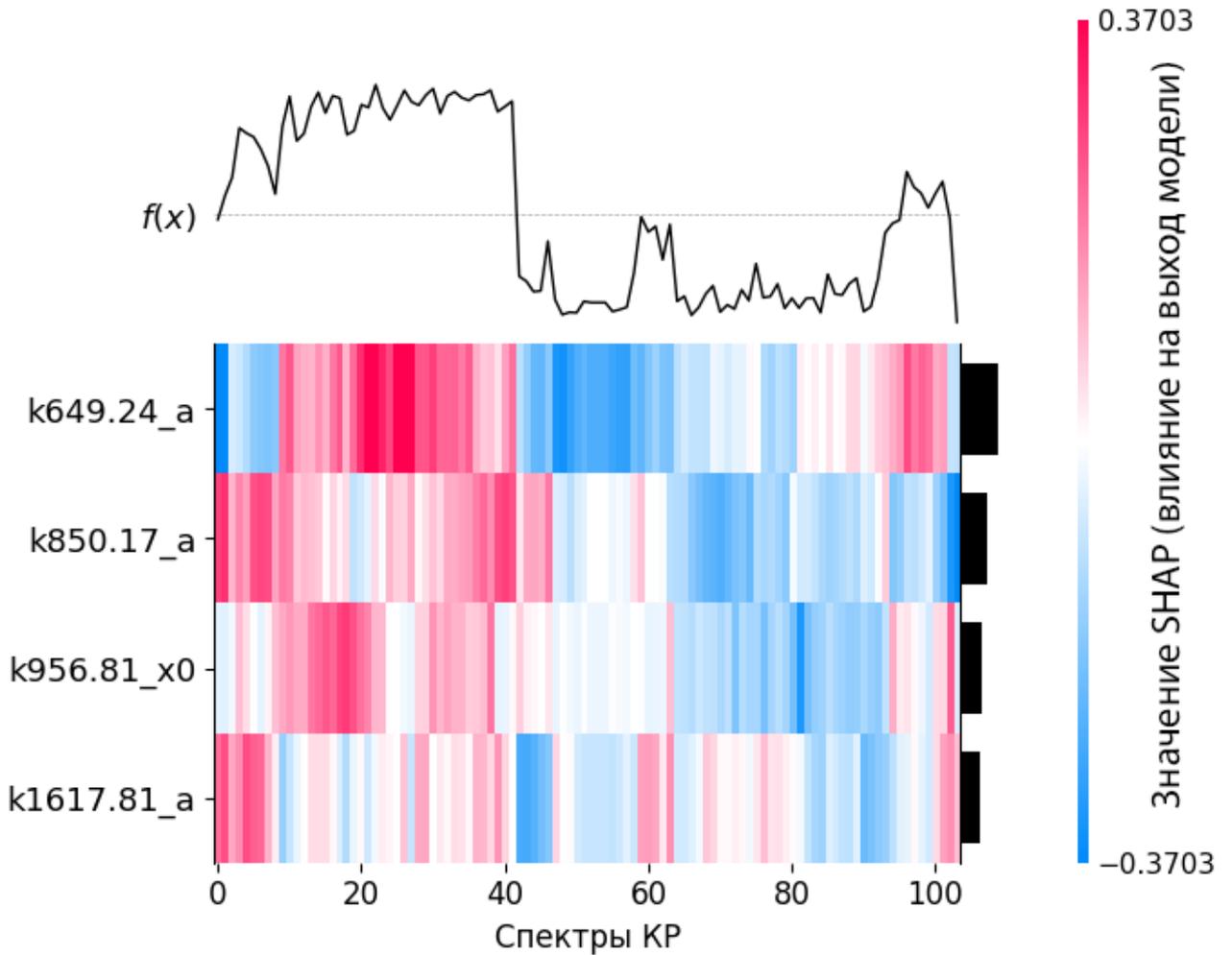


Рисунок 57 – Тепловая карта SHAP-значений. По оси Y отложены линии КР, упорядоченные по важности, по оси X – отдельные наблюдения (спектры). Красный цвет указывает на увеличение вероятности целевого класса, синий – на уменьшение. Значения SHAP нормированы в диапазоне от $-0,3703$ до $0,3703$.

Модель показывает высокое качество и интерпретируемость. Признаки адекватно отражают вклад в предсказание, наблюдается логичная и устойчивая картина SHAP-значений.

5.1.7 Оценка качества модели без использования декомпозиции

Для сравнения приведены значения метрик полученного классификатора (таблица 11), построенного на амплитудах на волновых числах ~ 649 , ~ 850 , ~ 957 и

$\sim 1618 \text{ см}^{-1}$. Положение максимума линий не анализировалось, так как без декомпозиции его выделить невозможно.

Таблица 11. Значения метрик модели логистической регрессии без использования декомпозиции спектров, доверительные интервалы рассчитаны для уровня вероятности не менее 95 %

Метрика	Обучающая выборка (87 спектров)	Тестовая выборка (29 спектров)
Точность (accuracy)	0,82 (0,72-0,90)	0,79 (0,62-0,93)
ROC AUC	0,91 (0,84-0,96)	0,81 (0,63-0,96)
Точность (precision)	0,85 (0,75-0,94)	0,83 (0,72-0,91)
Чувствительность (recall)	0,95 (0,81-0,99)	0,93 (0,79-0,96)
Специфичность	0,95 (0,81-0,99)	0,93 (0,79-0,96)
Оценка F1 (F-мера)	0,91 (0,85-0,94)	0,90 (0,84-0,93)
Логарифмическая потеря (Logloss)	0,36 (0,29-0,43)	0,40 (0,27-0,56)

Наблюдается резкое падение значений метрик по сравнению с моделью, построенной на амплитудах и положению максимума декомпозированных линий. В частности, ROC AUC снизился с 0,94 до 0,81. Что в первую очередь связано с отсутствием высокоинформативного признака соответствующего положению максимума линии $\sim 957 \text{ см}^{-1}$, а также со снижением информативности других линий (таблица 12) что приводит к недообучению модели.

Таблица 12. Результаты U-теста

Признак	<i>p</i> -значение (с декомпозицией)	<i>p</i> -значение (без декомпозиции)
k649.24	10^{-5}	0,028
k850.17	10^{-5}	0,28
k956.81	10^{-5}	10^{-5}
k1617.81	10^{-5}	0,17

5.1.8 Анализ *in vivo* спектров КР эмали для валидации полученной модели

Особое внимание заслуживает тот факт, что калиброванная модель выдаёт вероятности, что позволяет не только категорически различать объекты, но и определять вероятность наличия признаков пародонтита для каждого отдельного образца.

Для валидации разработанной модели классификации пародонтита были исследованы 40 спектров КР эмали добровольцев *in vivo* (диагноз на момент измерений не был известен). Ранее показано, что при пародонтите в спектрах эмали возникают характерные изменения, на основании которых можно построить неинвазивные диагностические критерии. В анализе использовалась обученная модель классификации с порогом вероятности 0,5 (считается что имеются признаки пародонтита если $p > 0,5$).

- Модель классифицировала 35 спектров (88 %) как пародонтит ($p > 0,5$).
- Из них 28 спектров (70 %) имели очень высокую вероятность ($p > 0,8$), т.е. высокая уверенность модели.
- Присвоены промежуточные вероятности 7 спектрам ($0,3 < p \leq 0,7$).
- Остальные 2 спектра (5 %) модель сочла здоровыми ($p \leq 0,3$).

Гистограмма распределения оценок вероятности пародонтита (рисунок 58) показывает резкое скопление значений около 0,9, при этом в области 0,5–0,8 наблюдается относительно малое число точек, а в низком диапазоне ($p \leq 0,5$) – лишь единичные спектры. Такой профиль отражает высокую уверенность модели в большинстве случаев и лишь редкие «пограничные» оценки.

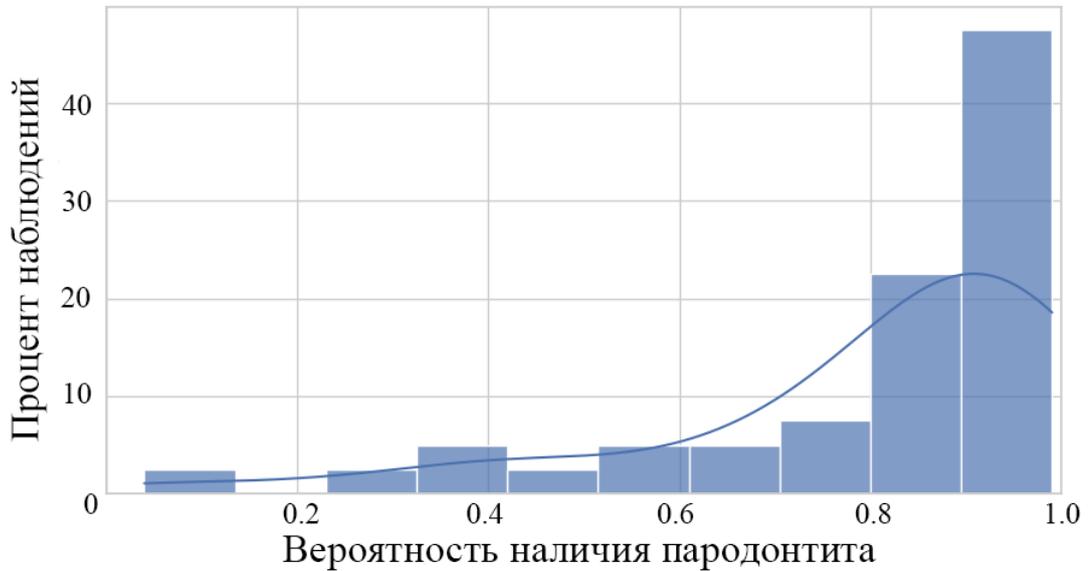


Рисунок 58 – Гистограмма распределения оценок вероятности

Кривая кумулятивной функции распределения вероятностей (рисунок 59) характеризуется стремительным ростом на участке $p > 0,8$ и практически выравнивается («плато») при значениях ниже примерно 0,4. Это означает, что свыше 80 % спектров дают предсказания с высокой уверенностью, тогда как остальные распределены по нижним уровням шкалы накопленной доли.

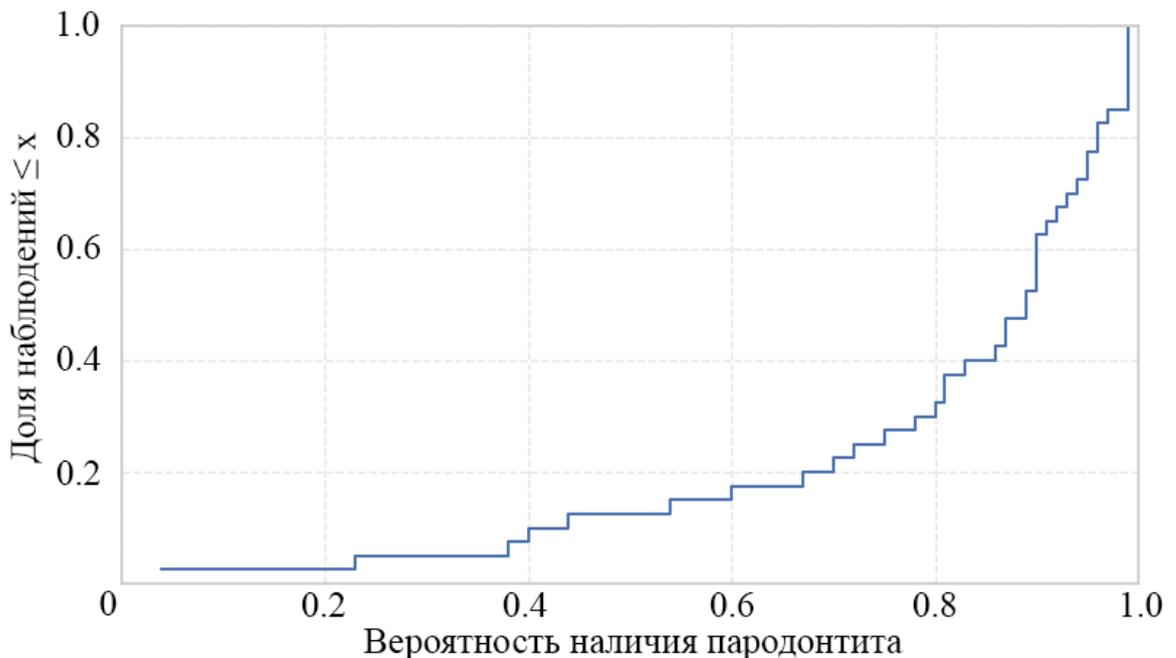


Рисунок 59 – Кривая кумулятивной функции распределения предсказанных вероятностей

Таким образом, модель выявила спектральные признаки пародонтита на доклинической стадии, задолго до появления рентгенологических изменений. У всех 5 добровольцев впоследствии были диагностированы пародонтит I–II стадии.

5.2 Применение спектроскопии КР для оценки состояния костной ткани при пародонтите

В разделе 5.2 показаны результаты исследований костной ткани зубов при пародонтите и в норме. Выявлены спектральные изменения костной ткани зубов. Установлено, что при пародонтите, в спектрах КР костной ткани происходят изменения интенсивностей линий КР компонентов органического матрикса и минеральной составляющей, что возможно является следствием резорбции костной ткани при данном заболевании.

Целью данной работы – оценка возможности использования предложенного способа анализа спектров КР с применением методов машинного обучения для диагностики признаков пародонтита в костной ткани.

5.2.1 Пробоподготовка и алгоритмы обработки спектров КР

В качестве объектов исследований была использована костная ткань пациентов, обоего пола, возрастной группы от 47 до 58 лет с установленным диагнозом пародонтит. Диагноз пародонтит был установлен клинически и при анализе компьютерной конусно-лучевой томографии (КТ) (код заболевания по МКБ-10 (1997 г.) - K05.3). Костная ткань в норме была получена при экстрагировании восьмых зубов.

Образцы были разделены на две группы: 1 группа – костная ткань в норме (26 спектров КР) и 2 группа – костная ткань при пародонтите (46 спектров КР).

Исследования проводились с помощью метода спектроскопии комбинационного рассеяния.

Полученные спектры комбинационного рассеяния далее обрабатывались алгоритмами в соответствии с блок-схемой на рисунке 60 до этапа получения таблицы с амплитудами декомпозированных линий КР.



Рисунок 60 – Блок схема математической обработки спектров КР

5.2.2 Анализ усредненных спектров КР и результаты декомпозиции

Рассмотрим нормированные усредненные спектры комбинационного рассеяния исследуемых образцов. Как видно из рисунка 61, спектры КР двух групп имеют ярко выраженные спектральные различия.

Для спектров КР костной ткани при пародонтите характерны более высокие значения относительной интенсивности линий чем для спектров контрольной

группы образцов: $\sim 1318 \text{ см}^{-1}$ (амид III или маркер гуанина) [134] [135], $\sim 1395 \text{ см}^{-1}$ и $\sim 1744 \text{ см}^{-1}$ (C=O колебания фосфолипидов) [136].

Для спектров КР контрольной группы характерны более высокие значения интенсивности по сравнению со спектрами КР костной ткани при пародонтите для линий: ~ 852 и ~ 917 (C–C колебания пролина), ~ 1039 ($\text{PO}_4^{3-} \nu_3$), ~ 1067 ($\text{CO}_3^{2-} \nu_1$) ~ 1244 (амид III), ~ 1444 (CH_2 колебания в протеинах и липидах) [139], и $\sim 1663 \text{ см}^{-1}$ (амид I).

Статистически значимых отличий в относительной интенсивности других линий КР не наблюдается. Коэффициент корреляции Пирсона между усредненными спектрами КР двух групп = 0,9636.

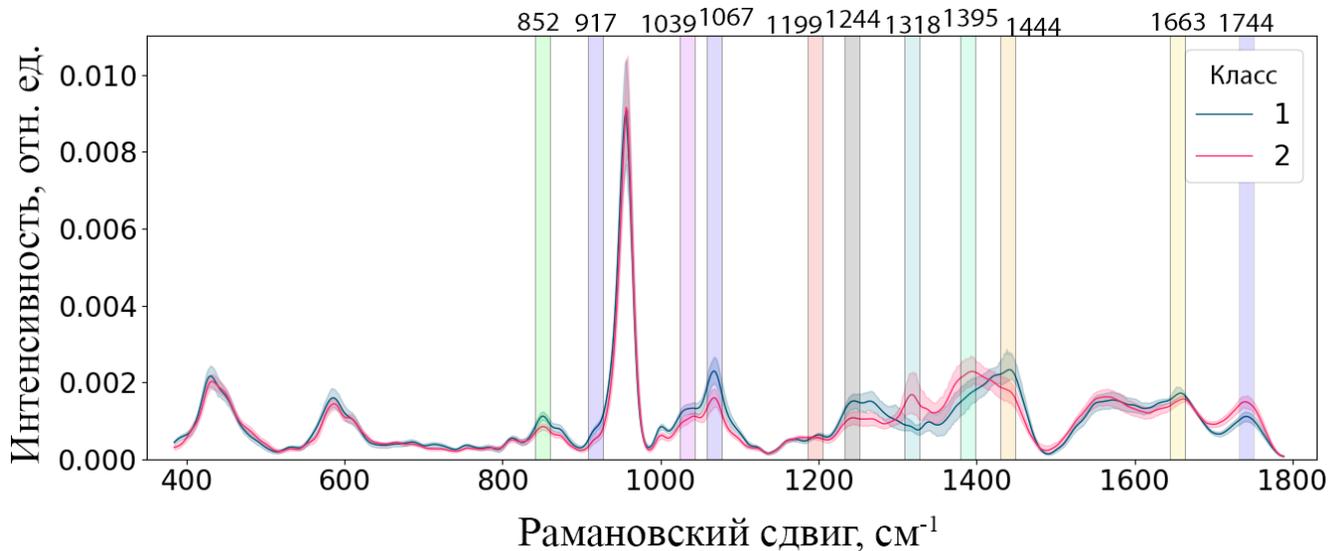


Рисунок 61 – Усредненные спектры КР исследуемых образцов: 1 – костная ткань при пародонтите, 2 – костная ткань в норме

Присутствие маркера ДНК гуанина, в частности маркеров В- и Z-формы, в области спектра 1318 см^{-1} может потенциально указывать на бактериальное участие в пародонтите. ДНК пародонтальных патогенов, таких как *Porphyromonas gingivalis* и других бактерий, участвующих в пародонтите, была изучена на предмет ее уникальных биомаркеров, которые могут вызывать иммунный ответ и способствовать прогрессированию заболевания.

Внеклеточная ДНК, высвобождаемая во время гибели бактериальных клеток, а также сама бактериальная ДНК играют важную роль в воспалительном процессе, связанном с пародонтитом. Эта ДНК может вызывать иммунные реакции,

приводящие к повреждению тканей и воспалению, типичному для пародонтоза. Исследования показали, что бактериальная ДНК и связанные с ней биомаркеры, такие как *P. gingivalis*, тесно связаны с тяжестью пародонтита. В частности, эпигенетические изменения, связанные с ДНК в тканях десен, включая изменения метилирования и экспрессии генов, могут еще больше усугубить воспалительную реакцию на эти бактерии [140] [141] [142].

Таким образом, различия, наблюдаемые при 1318 см^{-1} , могут быть связаны с наличием бактериальной ДНК, что потенциально указывает на повышенную бактериальную активность или иммунный ответ, оба из которых имеют решающее значение в развитии заболеваний пародонта. Необходимы дальнейшие исследования для полного изучения связи между этой спектральной особенностью и специфической бактериальной ДНК, связанной с пародонтитом.

Для декомпозиции спектров КР на 141 линии использован метод, описанный в 3 главе. Усредненное значение метрик для декомпозированных 72 спектров КР составило: $R^2 = 99,82 \%$, $\chi^2 = 2 \cdot 10^{-6}$, $\text{aic} = -16262$, $\text{bic} = -14267$.

В дальнейшем статистическом анализе использовались амплитуды декомпозированных линий a .

5.2.3 Отбор признаков

На первом этапе отбора признаков амплитуды декомпозированных линий были проанализированы методом ANOVA (Дисперсионный анализ с использованием класса `SelectPercentile` библиотеки `sklearn` с параметром `percentile = 50`) и из 141 было отобрано 70 признаков.

На втором этапе отбора признаков была применена встроенная функция `select_features` градиентного бустинга на решающих деревьях `CatBoostClassifier`, ориентированная на итеративное исключение признаков с наименьшим вкладом в предсказательную способность модели. Метод реализует разновидность рекурсивного исключения признаков с использованием собственной метрики качества модели.

Алгоритм последовательно исключал признаки с наименьшим вкладом, производя оценку качества модели на отложенной валидационной выборке (20 % от исходных данных, стратифицированной по целевой переменной). В результате данного этапа был сформирован поднабор из 25 признаков, обладающих наибольшей значимостью с точки зрения повышения обобщающей способности модели.

На заключительном этапе отбора признаков был проведён корреляционный анализ с целью исключения избыточных (высококоррелированных) признаков, способных вносить мультиколлинеарность в модель (рисунок 62).

Дополнительно на данном этапе были внесены экспертные корректировки на основе предварительных результатов моделирования. В результате итоговый набор был сокращён до 5 информативных и взаимно слабо коррелированных признаков, представляющих наибольший интерес как с точки зрения статистической значимости, так и с точки зрения интерпретируемости модели., соответствующих амплитуде декомпозированных линий на ~ 852 , ~ 1199 , ~ 1318 , ~ 1663 и ~ 1744 см^{-1} .

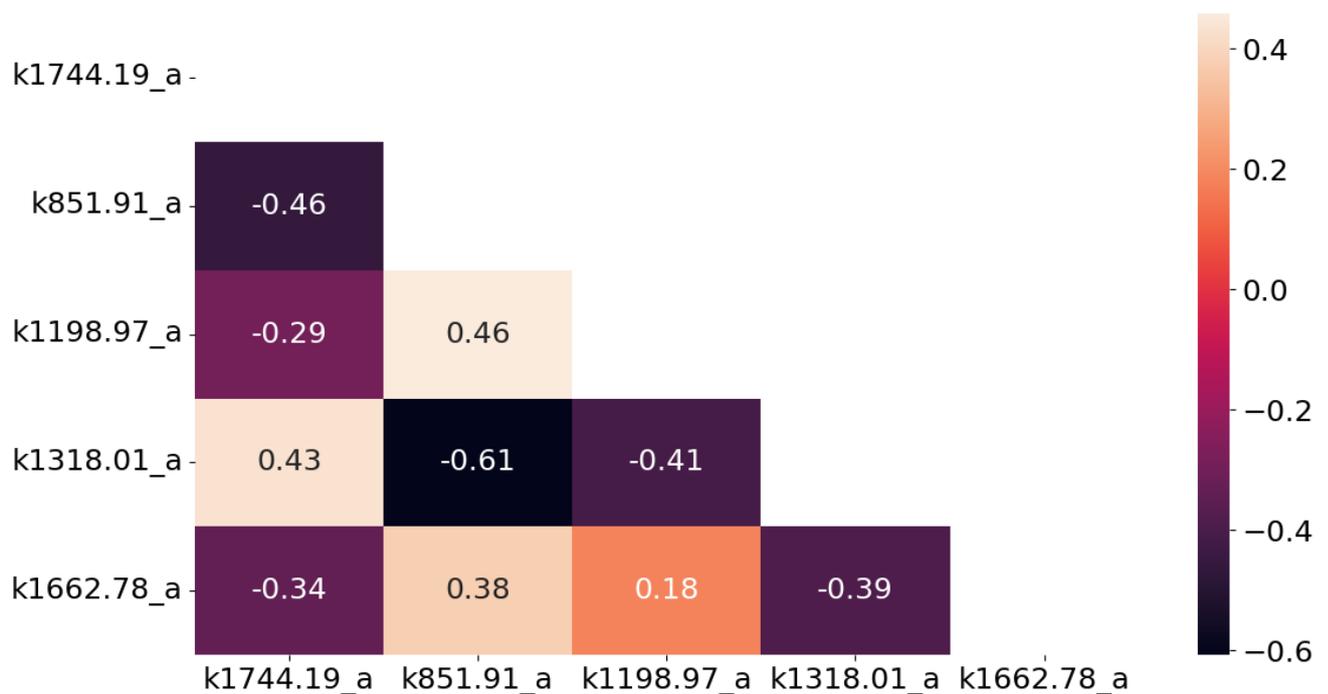


Рисунок 62 – Корреляционная матрица отобранных признаков

5.2.4 Статистический анализ различий между группами

5.2.4.1 Проверка нормальности распределения

Перед проведением непараметрических тестов была проведена проверка нормальности распределения признаков в каждой из сравниваемых групп с использованием критерия Шапиро–Уилка. Полученные p -значения:

Для группы 1 (контрольная группа): $3 \cdot 10^{-7}$;

Для группы 2 (эмаль с диагнозом пародонтит): 0,003.

Обе группы демонстрируют значимые отклонения от нормального распределения (p -значение $< 0,01$), что оправдывает использование непараметрических методов сравнения, в частности – пермутационного теста и критерия Манна–Уитни.

5.2.4.2 Пермутационный тест

Результаты теста:

- Статистика теста: 0,00059;
- p -значение: 0,00081;
- Расчетная мощность теста: 0,99;
- Необходимый минимальный размер выборки для достижения мощности 0,8: 8 спектров на группу.

Результаты свидетельствует о высокой чувствительности теста при текущем размере выборки и значительном эффекте между группами.

5.2.4.3 Критерий Манна–Уитни (U-тест)

U-тест был применён к каждому признаку по отдельности. Для каждой переменной рассчитывались: статистика U и соответствующее p -значение.

В таблице 13 представлены значения U-статистики и соответствующих уровней значимости (p -значение) для каждого из 6 признаков.

Таблица 13. Результаты U-теста

Признак	U-статистика	<i>p</i> -значение
k851.91_a	923,0	0,00014
k1198.97_a	924,5	0,00013
k1318.01_a	309,5	0,00073
k1662.78_a	803,0	0,016
k1744.19_a	237,0	0,000024

Во всех случаях наблюдаются значения *p*-значения меньше 0,01 (кроме признака k1662.78_a), что свидетельствует о наличии статистически значимых различий между группами по каждому из исследуемых признаков.

Дополнительно результаты U-теста были подтверждены с использованием пермутационного теста, что позволило исключить влияние случайных факторов и повысить доверие к полученным результатам. Совпадение результатов обоих методов и наличие устойчиво низких *p*-значений указывает на высокую статистическую мощность анализа и подтверждает, что объём выборки является достаточным для выявления существующих эффектов.

5.1.4.4 Распределения значений признаков

Рисунок 63 демонстрирует распределения амплитуд наиболее информативных линий КР, отобранных на основании результатов статистического анализа. Графики построены отдельно для контрольной группы (отмечена синим цветом) и группы пациентов с диагнозом пародонтит (отмечена красным цветом). Наблюдается отчётливая разница в распределениях, что указывает на возможные структурные и/или биохимические изменения в составе костной ткани.

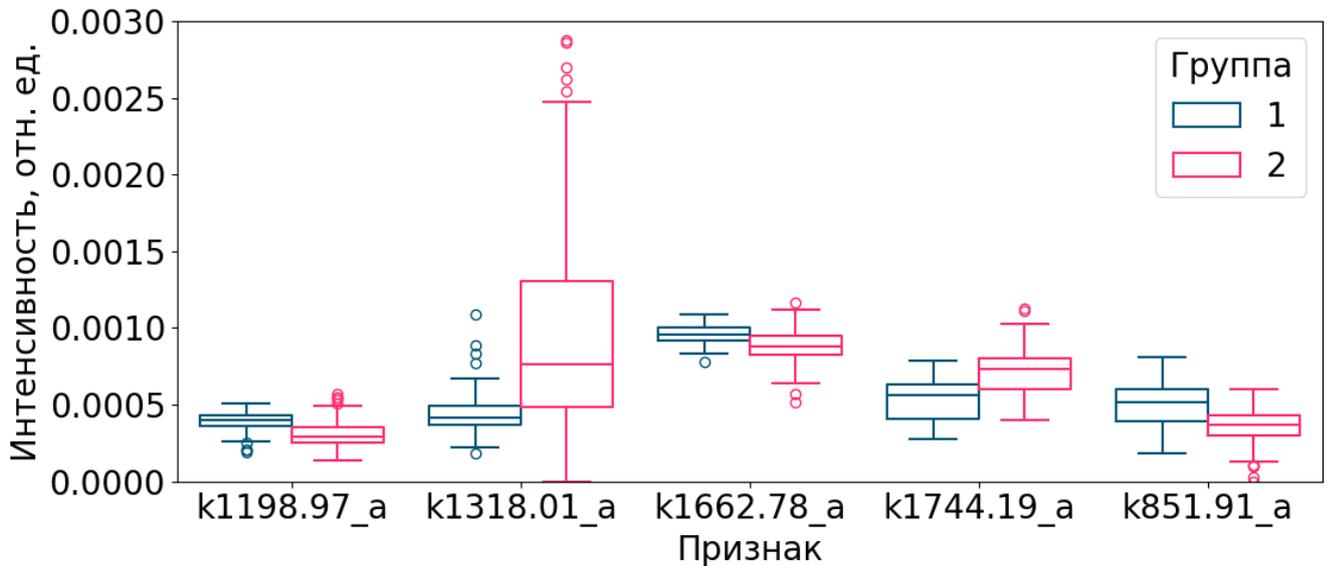


Рисунок 63 – Распределения амплитуды наиболее информативных линии КР: 1 – контрольная группа, 2 – с диагнозом пародонтит

5.2.5 Процесс создания модели машинного обучения

В рамках данного исследования в качестве базовой модели машинного обучения была выбрана логистическая регрессия. Выбор данной обоснован ранее в разделе 4.1.5.1.

5.2.5.1 Подготовка выборки для обучения модели

Для повышения объема обучающей выборки, увеличения обобщающей способности и обеспечения устойчивости моделей машинного обучения к ограниченности данных была реализована процедура синтетической генерации данных с применением модели Conditional Tabular GAN (CTGAN - генеративно-состязательная сеть для работы с табличными данными). Генерация проводилась отдельно для каждого класса с учетом специфики признаков. Выполнялось масштабирование числовых признаков в диапазон $[0, 1]$.

Для каждого класса и каждого признака проводилось визуальное сравнение распределений между оригинальными и синтетическими данными. Это позволило выявить отклонения и визуально оценить адекватность сгенерированной выборки.

Итоговое количество спектров составило:

- Группа 1 (контрольная группа): 78 спектров,
- Группа 2 (костная ткань с диагнозом пародонтит): 138 спектров.

Несмотря на некоторое числовое различие между группами, дисбаланс классов является незначительным, что обеспечивает корректное представление распределения классов как в обучающей, так и в тестовой выборках.

Выборка разделялась на обучающую и тестовую случайным образом в отношении 75:25 с использованием стратификации по целевому признаку.

Перед обучением модели данные подвергаются дополнительной обработке, направленной на улучшение качества обучения. В рамках конвейера обработки используется `QuantileTransformer`.

5.2.5.2 Обучение модели логистической регрессии

Для построения базовой модели бинарной классификации использовалась логистическая регрессия, реализованная в рамках конвейера (`Pipeline`), включающего этап предварительной обработки признаков с использованием `QuantileTransformer` и калибровки модели с помощью сигмоиды.

Обучение модели сопровождалось оптимизацией гиперпараметров методом байесовской оптимизации с помощью библиотеки `Optuna`. Целевой функцией являлось среднее значение площади под ROC-кривой (ROC AUC), рассчитанное на основе стратифицированной кросс-валидации с 3 фолдами и Оценка Брайера (Brier-Score). Такой подход обеспечивает устойчивую оценку обобщающей способности модели, калибровки и снижает риск переобучения, особенно при ограниченном объеме данных.

Найденные оптимальные значения приведены в таблице 14.

Таблица 14. Гиперпараметры модели и их значения

Параметр	Значение	Обоснование
penalty	12	L2-регуляризация (ридж-регуляризация) штрафует большие веса, уменьшая переобучение
C	0,28	Умеренная регуляризация. Наилучший баланс между недообучением (при слишком сильной регуляризации) и переобучением (при слишком слабой)
Solver_l2	newton-cholesky	Метод Ньютона с факторизацией Холецкого, быстро сходится на небольших и средних датасетах
max_iter	5000	Увеличенное количество итераций обеспечивает сходимость при трансформированных данных
class_weight	'balanced'	Автоматически компенсирует незначительный дисбаланс классов. Особенно важно при использовании AUC как метрики.

5.2.6 Оценка качества и интерпретация полученного классификатора

Модель оценивает состояние костной ткани при пародонтите по пяти ключевым спектральным признакам (амплитуды декомпозированных линий при ~ 852 , ~ 1199 , ~ 1318 , ~ 1663 и ~ 1744 см^{-1}), которые после квантильного преобразования к нормальному распределению подаются на вход логистической регрессии, которая выдаёт откалиброванную вероятность p наличия патологических изменений, причём для бинарного решения обычно используется порог $p > 0.5$: образцы с $p \leq 0,5$ считаются нормой, а с $p > 0,5$ – признаком пародонтита.

Таблица 15 содержит основные метрики качества модели, полученной на выборках, состоящих из 162 спектров для обучения и 54 спектров для валидации. Значения точности (accuracy) составляют 0,85 для обучающей выборки и 0,80 для тестовой. Показатель ROC AUC, равный 0,92 и 0,88 соответственно, подтверждает высокую чувствительность и специфичность модели. Анализ значений precision,

recall и F1-меры также позволяет отметить, что модель обучена стабильно, не переобучена, хорошо обобщается на новые данные и даёт уверенные предсказания. Таблица 15. Значения метрик модели логистической регрессии, доверительные интервалы рассчитаны для уровня вероятности не менее 95 %

Метрика	Обучающая выборка	Тестовая выборка
Точность (accuracy)	0,85 (0,80-0,90)	0,80 (0,69-0,91)
ROC AUC	0,92 (0,87-0,96)	0,88 (0,78-0,96)
Точность (precision)	0,86 (0,75-0,95)	0,68 (0,47-0,87)
Чувствительность (recall)	0,79 (0,59-0,95)	0,71 (0,59-0,82)
Специфичность	0,79 (0,59-0,95)	0,71 (0,59-0,82)
Оценка F1 (F-мера)	0,77 (0,68-0,86)	0,72 (0,55-0,86)
Логарифмическая потеря (Logloss)	0,38 (0,33-0,44)	0,43 (0,35-0,52)

Распределение скоров модели (score) (рисунок 64) демонстрирует определённое перекрытие, однако наблюдаемое смещение средних значений указывает на способность модели учитывать различия между классами. Этот результат подтверждает, что, несмотря на перекрытие, входные признаки обладают диагностической информацией, позволяющей классификатору выделять патологические состояния.

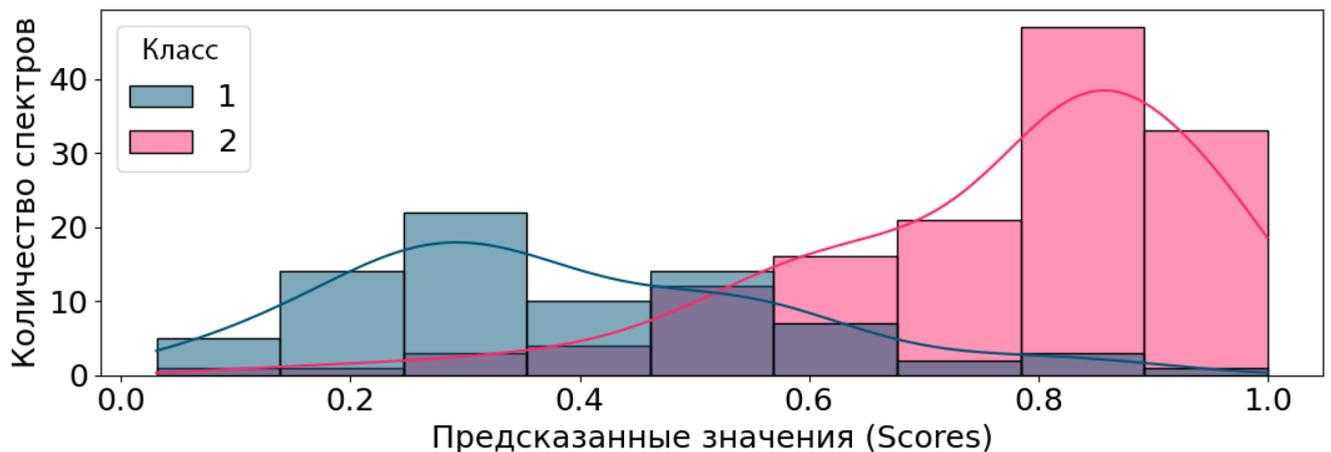


Рисунок 64 – Выходные значения модели (score). 1 – контрольная группа, 2 – с диагнозом пародонтит

На проекции с использованием МГК (рисунок 65) наблюдается пересечение классов. Видно, что первая главная компонента (PC-1) объясняет 93,63 %

дисперсии, что указывает на практическую одномерность данных. Наблюдается, что логистическая регрессия формирует линейно разделяющее пространство, соответствующее характеру признаков. Наличие перекрытия между классами в области около 0,5 по РС-1 указывает на частичную неразделимость классов и возможное присутствие шумовых наблюдений.

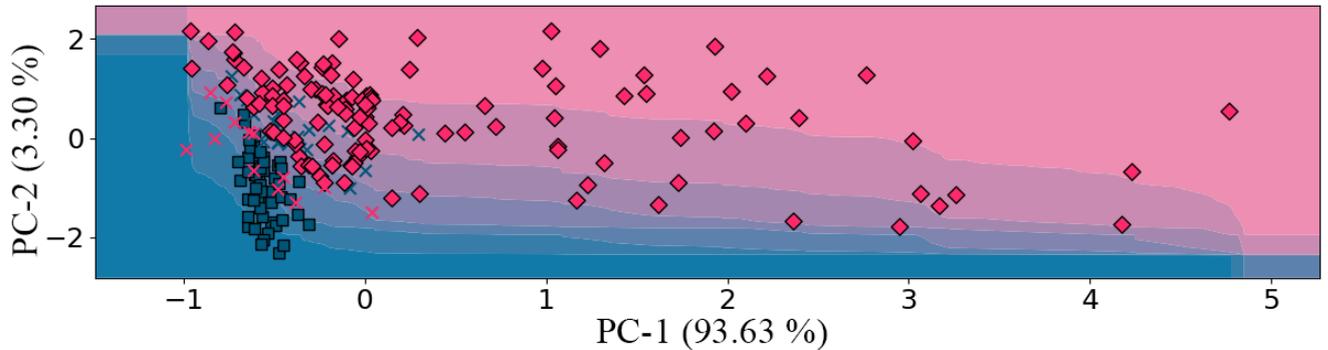


Рисунок 65 – Визуализация границ принятия решения модели логистической регрессии в пространстве главных компонент. Синий – контрольная группа, красный – с диагнозом пародонтит

Матрица ошибок (рисунок 66) подтверждает, что классификатор обладает приемлемой способностью к различению классов, демонстрируя устойчивость и сбалансированную производительность. Это подтверждает ранее рассчитанные метрики (accuracy, precision, recall, F1).

Истинный класс	1	15	4
	2	7	28
		1	2
		Предсказанный класс	

Рисунок 66 – Матрица решений

Кривая ошибок первого и второго рода (ROC AUC curve) показана на рисунке 67, площадь под кривой (AUC) составляет 0,88, что говорит о хорошей классифицирующей способности модели.

Анализ кривой точности – полноты (Precision-Recall curve) (рисунок 67) показал, что логистическая регрессия достигает высокого значения средней

точности (Average Precision) (AP = 0,94), что свидетельствует о низком количестве ошибок классификации.

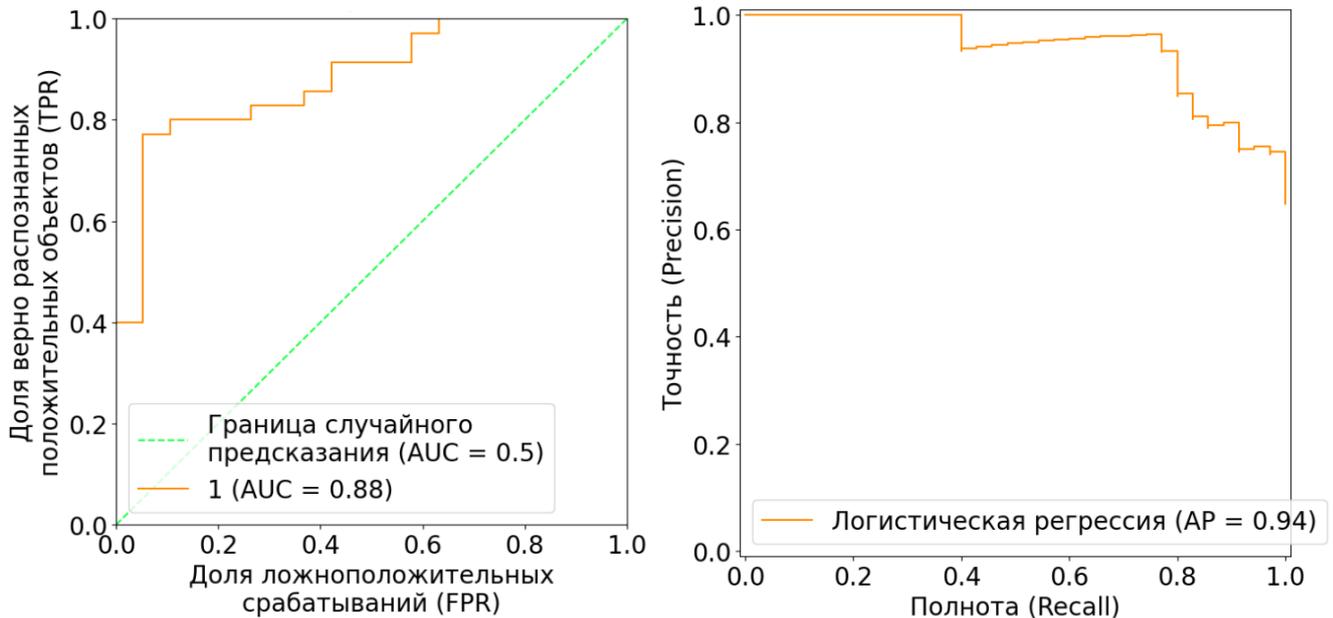


Рисунок 67 – Кривая ошибок первого и второго рода (ROC AUC curve) (слева) и кривая точности – полноты (Precision – Recall curve) (справа)

Для интерпретации вклада отдельных признаков в предсказания модели логистической регрессии был построен график значений SHAP (рисунок 68), отражающий влияние каждого признака на выход модели.

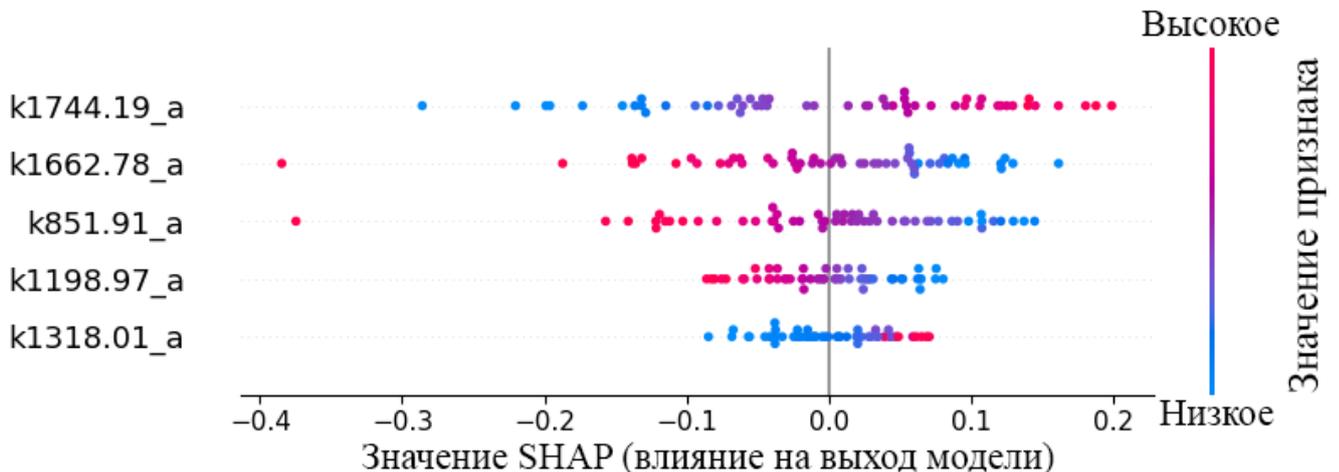


Рисунок 68 – график SHAP-значений для логистической регрессии. По оси Y отложены линии КР, по оси X – величина их вклада в предсказание модели (SHAP). Красный цвет соответствует высоким значениям признака, синий – низким. Положительные SHAP-значения увеличивают вероятность прогноза как целевого класса (пародонтит), отрицательные – уменьшают

Суммарно, видно, что признаками пародонтита в костной ткани являются высокие значения признаков, соответствующих амплитуде линий 1318 и 1744 см^{-1} . И соответственно более низкие значения остальных признаков.

На тепловой карте (рисунок 69) наблюдения (спектры по оси X) упорядочены по значению функции модели $f(x)$, то есть по выходу модели. Это видно из плавного графика над тепловой картой, который отражает последовательность предсказаний – от высоких значений к низким (или наоборот), позволяя сопоставить изменение SHAP-вкладов с изменением предсказания.

Логика модели устойчива. Наблюдается систематическое поведение SHAP-значений вдоль оси X – например, на высоких значениях $f(x)$ преобладает красный цвет, а на низких – синий. Это означает, что влияние признаков согласуется с предсказанием модели, и она ведёт себя логично.

Поскольку модель уверенно усиливает или подавляет вероятность предсказания с помощью разных признаков (в зависимости от спектра), можно заключить, что она обладает дискриминационной способностью и выявляет паттерны, характерные для целевого класса.

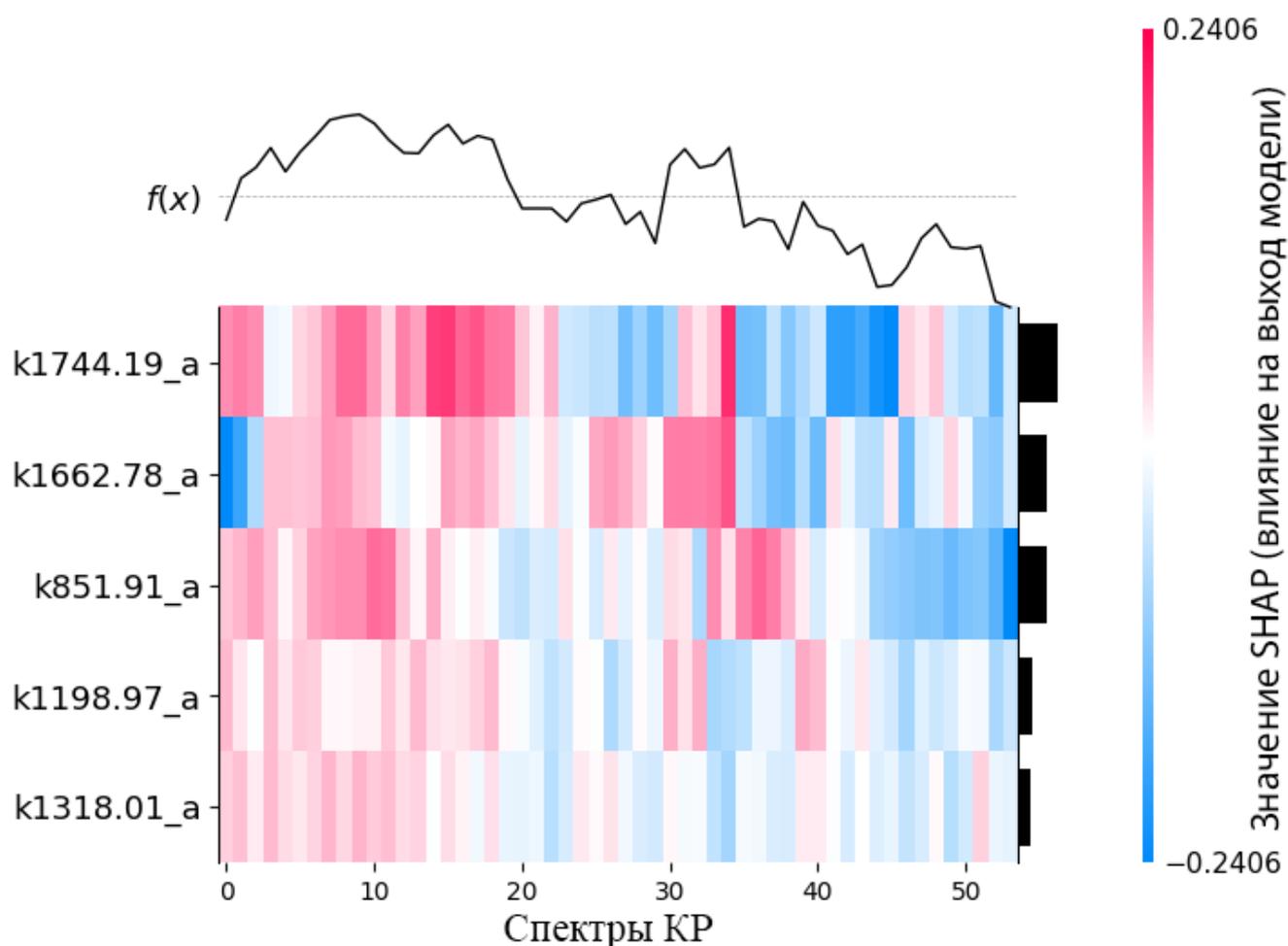


Рисунок 69 – Тепловая карта SHAP-значений. По оси Y отложены линии КР, упорядоченные по важности, по оси X – отдельные наблюдения (спектры). Красный цвет указывает на увеличение вероятности целевого класса, синий – на уменьшение. Значения SHAP нормированы в диапазоне от $-0,2406$ до $0,2406$.

Модель показывает высокое качество и интерпретируемость. Признаки адекватно отражают вклад в предсказание, наблюдается логичная и устойчивая картина SHAP-значений.

5.3 Применение спектроскопии КР для оценки твердых тканей зубов

Одним из важнейших этапов лечения является профессиональная гигиена полости рта, включающая удаление над- и поддесневого налета, снятие зубного камня и пломбирование корней. В ряде случаев для повышения эффективности лечение корня дополнялось рассечением внутренней эпителиальной выстилки пародонтального кармана (кюретаж), что является хирургической манипуляцией с

более длительным сроком заживления и непредсказуемым восстановлением прикрепления десны [143] [144] [145].

Необходимость и эффективность закрытого выскабливания по сравнению с профессиональной гигиеной полости рта все еще обсуждается. Причина в том, что изменения в тканях зуба после процедур еще до конца не изучены. В литературе существуют противоположные мнения о необходимости рассечения внутренней эпителиальной выстилки пародонтального кармана в процессе лечения корней, поэтому вопрос требует дальнейшего изучения [143] [146]. Полученные результаты будут способствовать поиску дополнительных методов лечения, которые можно было бы включить в комплексную терапию воспалительных заболеваний пародонта.

Целью данной работы было использование спектроскопии КР для оценки твердых тканей зубов после профессиональной гигиены полости рта и кюретажа.

5.3.1 Пробоподготовка и алгоритмы обработки спектров КР

В качестве объектов исследований было использовано 12 удаленных зубов пациентов, обоего пола. Удаление зубов проводилось в связи с хроническим пародонтитом средней степени, с глубиной пародонтальных карманов до 5 мм. Диагноз пародонтит был установлен клинически и при анализе компьютерной конусно-лучевой томографии (КТ) (код заболевания по МКБ-10 (1997 г.) - K05.3).

Зубы были исследованы до и спустя 6 недель после операции открытого кюретажа. Все снятые спектры КР были поделены на 2 основные группы: I группа (12 зубов) – зубы с диагнозом хронический пародонтит до стоматологических вмешательств, II группа (12 зубов) – те же самые зубы, с диагнозом хронический пародонтит, удаленные через 6 недель после закрытого кюретажа. Образцы для I и II групп брали у одних и тех же пациентов, с одной стороны челюсти осуществляя обработку корней по типу профессиональной гигиены полости рта, включающей удаление над- и поддесневых отложений, шлифование и полирование корней (Scaling и Root Planing), а с другой дополнительно с обработкой корня иссекали

эпителиальную выстилку пародонтального кармана, тем самым осуществляя закрытый кюретаж.

Исследовались поверхности зубов, в 3-х различных его областях: эмаль, дентин (в продольных срезах) и цемент. В каждой исследуемой области были сняты спектры в различных точках поверхности каждой тканей зуба. Всего исследовалось спектров КР: 86 эмали, 44 дентина и 66 цемента зубов.

Полученные спектры комбинационного рассеяния далее обрабатывались алгоритмами в соответствии с блок-схемой на рисунке 70 до этапа получения таблицы с амплитудами декомпозированных линий КР.

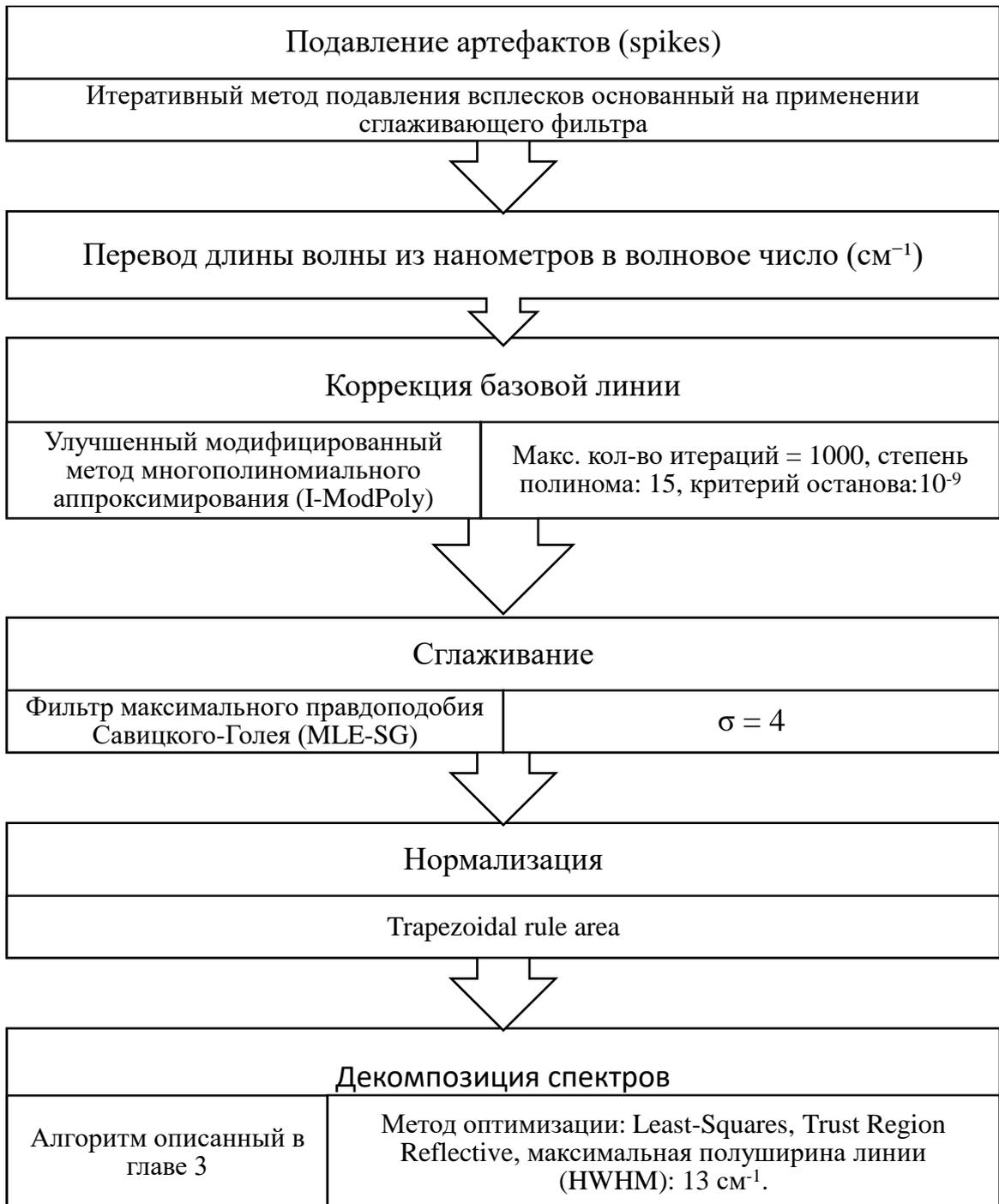


Рисунок 70 – Блок схема математической обработки спектров КР

5.3.2 Анализ усредненных спектров КР

Рассмотрим нормированные усредненные спектры комбинационного рассеяния эмали, дентина и цемента зубов (рисунок 71, 72, 73).

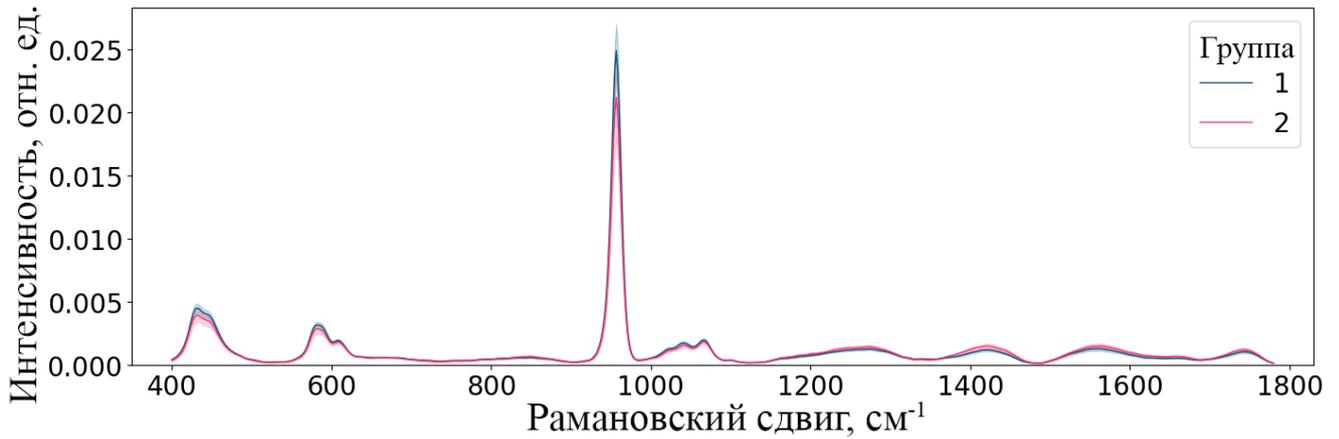


Рисунок 71 – Нормированные спектры КР эмали зубов исследуемых групп 1: зубы до стоматологических вмешательств, 2: зубы удаленные через 6 недель после закрытого кюретажа

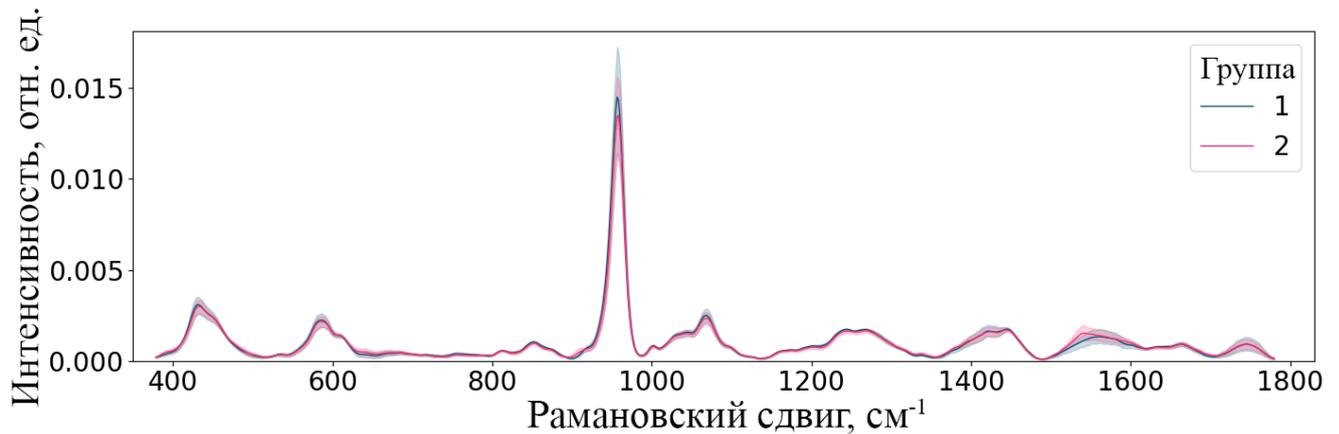


Рисунок 72 – Нормированные спектры КР дентина зубов исследуемых групп 1: зубы с диагнозом хронический пародонтит до стоматологических вмешательств, 2: зубы удаленные через 6 недель после закрытого кюретажа

Анализ спектров КР различных тканей зубов всех исследуемых групп показывает, что после профессиональной гигиены полости рта и процедуры закрытого кюретажа происходят схожие спектральные изменения (усиление интенсивности на линиях $\sim 589 \text{ см}^{-1}$, $\sim 432 \text{ см}^{-1}$, $\sim 957 \text{ см}^{-1}$, соответствующих гидроксиапатиту и $\sim 1068 \text{ см}^{-1}$, соответствующая С-О плоскостному валентному колебанию карбонат-иона гидроксиапатита $\text{CO}_3^{2-\nu_1}$) по сравнению с группой до проведения данных манипуляций. Данные спектральные изменения наиболее ярко проявляются в дентине, а также в цементе зубов.

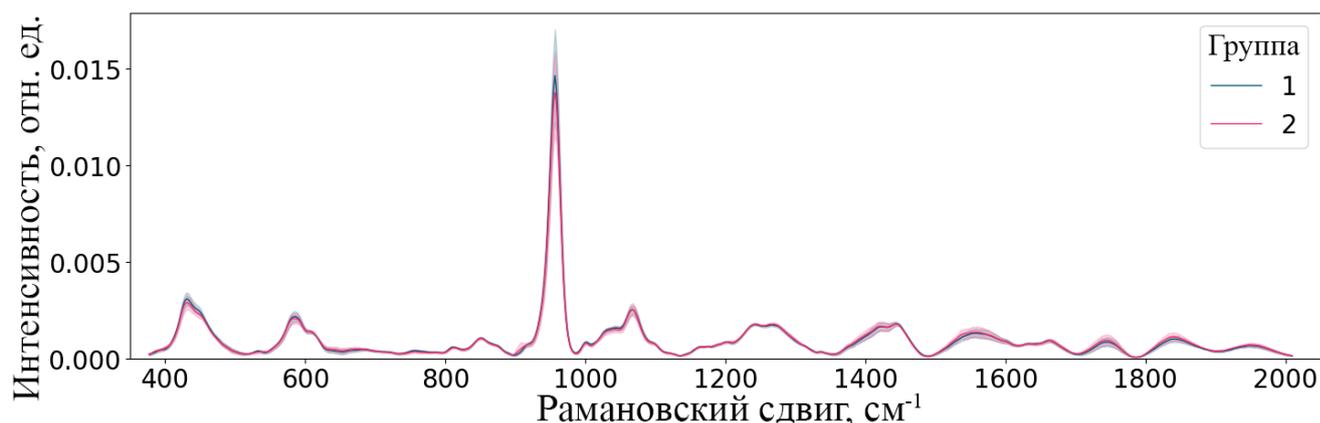


Рисунок 73 – Нормированные спектры КР цемента зубов исследуемых групп. 1: зубы с диагнозом хронический пародонтит до стоматологических вмешательств, 2: зубы удаленные через 6 недель после закрытого кюретажа

Возможные спектральные изменения в дентине зубов после обработки корней обусловлены проникновением активных минеральных компонентов в более глубокие ткани (дентин) зубов через специальные дентинные каналы. Спектральные изменения цемента зубов, после данных манипуляций, обусловлены восстановлением минеральных компонентов в тканях и образованием нового цемента за счет удаления зубного камня и процессов реминерализации [147].

Коэффициент корреляции Пирсона между усредненными спектрами эмали составляет 0,998, дентина – 0,998, цемента – 0,999.

Для декомпозиции спектров КР использован алгоритм, описанный в 3 главе.

5.3.3 Статистический анализ различий между группами

5.3.3.1 Проверка нормальности распределения

Перед проведением непараметрических тестов была проведена проверка нормальности распределения признаков в каждой из сравниваемых групп с использованием критерия Шапиро–Уилка. Полученные *p*-значения:

Для группы 1 и 2 спектров КР эмали: 0,0014 и 0,00073;

Для спектров КР дентина: 0,0051 и 0,39;

Для спектров КР цемента: $3 \cdot 10^{-6}$ и $8 \cdot 10^{-8}$.

Обе группы демонстрируют значимые отклонения от нормального распределения (*p*-значение < 0.01), что оправдывает использование

непараметрических тестов сравнения, в частности – пермутационного теста (таблица 16).

5.2.3.2 Пермутационный тест

Таблица 16. Результаты перестановочного теста:

Образцы	Статистика теста	p -значение	Расчетная мощность теста	Необходимый минимальный размер выборки для достижения мощности 0,8
Эмаль	0,0036	0,016	1,0	7
Дентин	0,0004	0,039	0,99	8
Цемент	0,00016	0,012	0,99	11

Результаты свидетельствует о высокой чувствительности теста при текущем размере выборки и отсутствии значимых различий между группами (p -значение $> 0,01$).

5.3.3.3 Анализ методом главных компонент

На рисунках 74, 75, 76 представлен результат анализа данных МГК. Анализ проводился с целью визуализации и интерпретации спектральных данных, полученных после декомпозиции спектров комбинационного рассеяния (КР) для сравнения спектров двух групп. При переходе в пространство первых двух главных компонент распределения спектров групп продемонстрировали значительное перекрытие, что указывает на высокую степень схожести значений относительных интенсивностей.

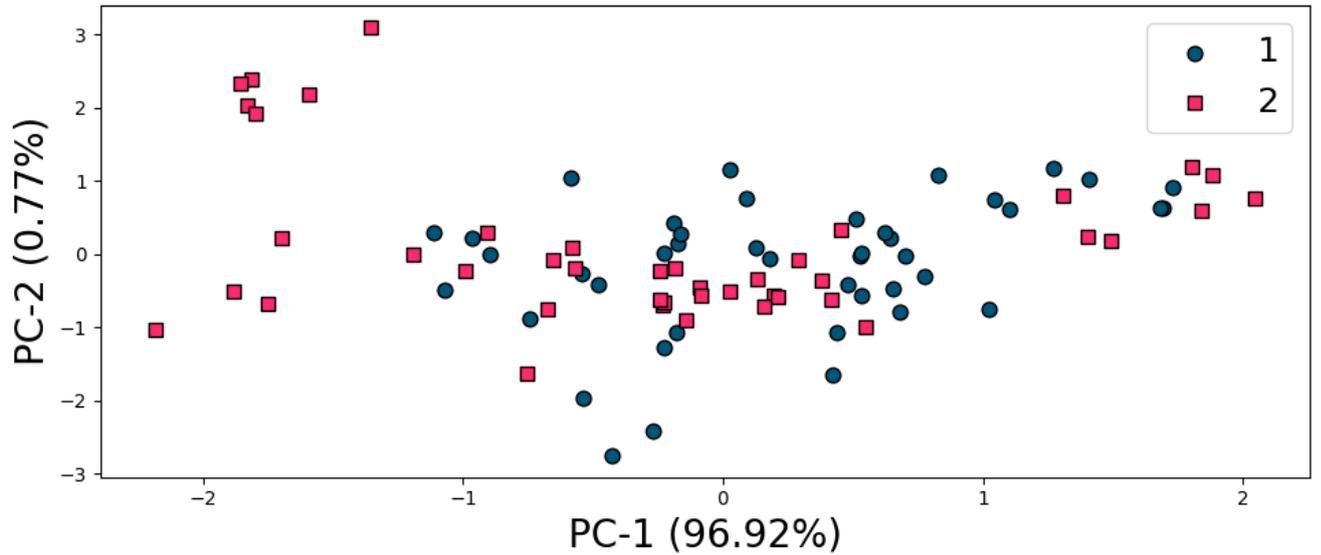


Рисунок 74 – График решений модели для данных в пониженной размерности для линий КР эмали. 1: зубы с диагнозом хронический пародонтит до стоматологических вмешательств, 2: зубы удаленные через 6 недель после закрытого кюретажа

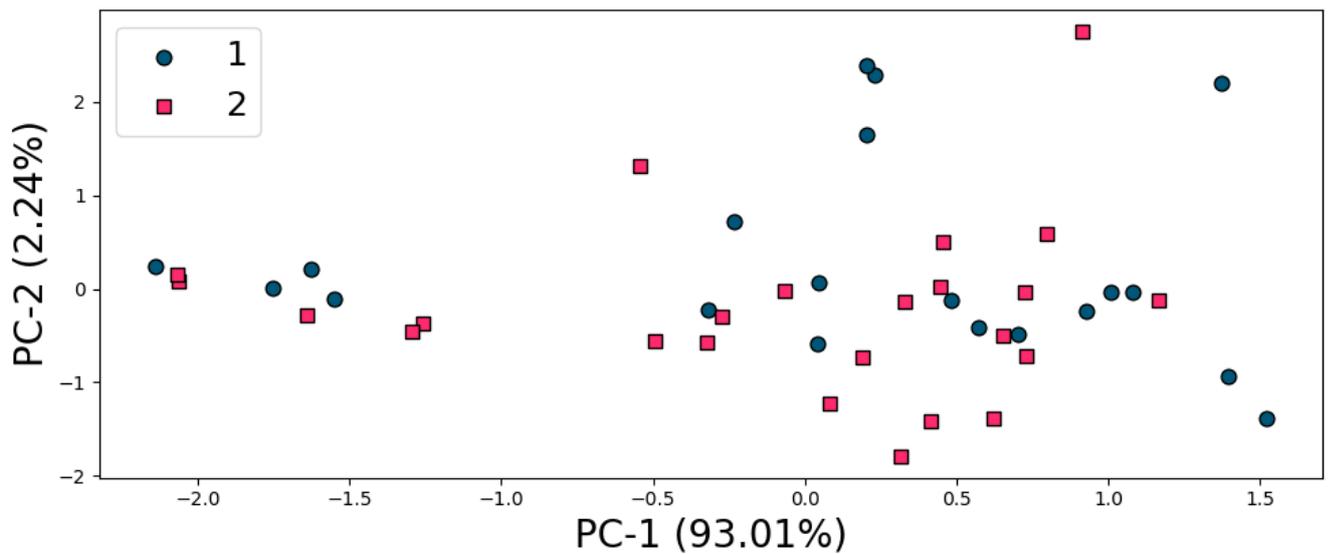


Рисунок 75 – График решений модели для данных в пониженной размерности для линий КР дентина.

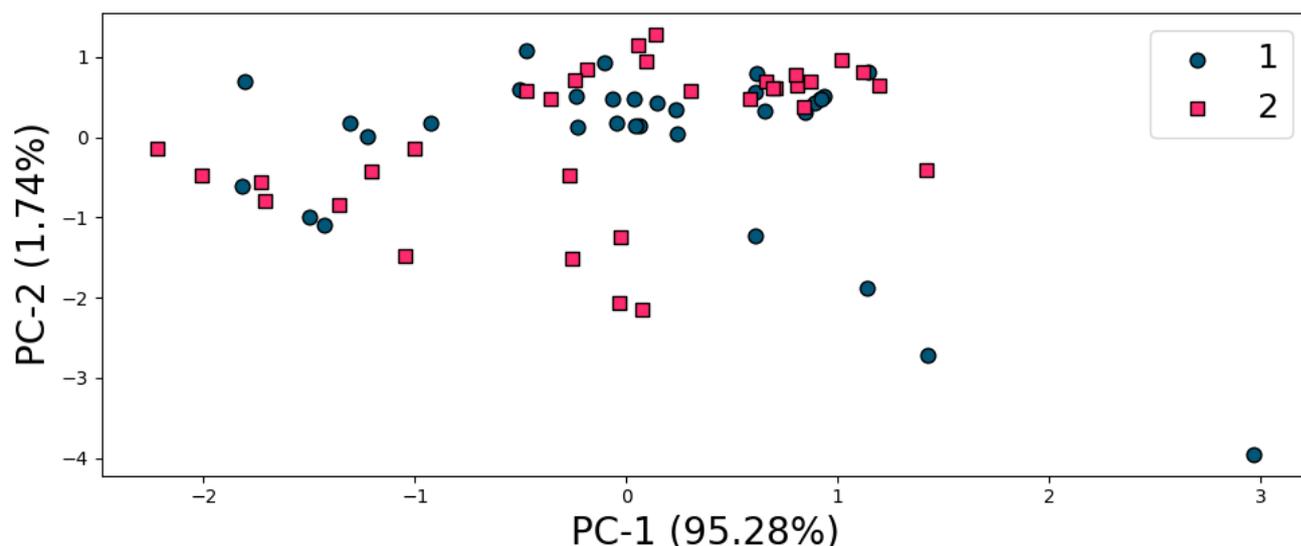


Рисунок 76 – График решений модели для данных в пониженной размерности для линий КР цемента.

5.4 Выводы по главе

В ходе исследования были получены данные, подтверждающие возможность применения разработанных методов для диагностики пародонтита и мониторинга процесса лечения. Это открывает перспективы для создания экспресс-методов, которые могут быть использованы в стоматологической практике для ранней диагностики и контроля лечения пародонтита.

1. Была поставлена и успешно реализована задача разработки неинвазивного способа экспресс-диагностики пародонтита на основе спектроскопии комбинационного рассеяния (КР) и логистической регрессии. Экспериментальная база включала 415 спектров КР эмали зубов, полученных как *in vitro*, так и *in vivo*, что обеспечило широкую репрезентативность. Проведена декомпозиция спектров с применением МНК ($R^2 = 99,92\%$, $\chi^2 = 2 \times 10^{-6}$, $AIC = -15685$, $BIC = -13681$), позволившая выделить ключевые спектральные линии, чувствительные к патологическим изменениям (амплитуда линий ~ 649 , ~ 850 и $\sim 1618\text{ см}^{-1}$ и положение максимума $\sim 957\text{ см}^{-1}$). Пермутационный тест и критерий Манна–Уитни показали статистически значимые различия по всем признакам (p -значение $< 0,01$), что подтверждает достоверность выявленных эффектов.

Разработанная модель логистической регрессии с использованием генерации синтетических данных (CTGAN) продемонстрировала высокие значения метрик на тестовой выборке: точность – 0,88 (0,82-0,94, 0,95 ДИ), ROC AUC – 0,94 (0,89-0,98, 0,95 ДИ), F1-мера – 0,89 (0,82-0,94, 0,95 ДИ), чувствительность – 0,86 (0,76-0,94, 0,95 ДИ), специфичность – 0,86 (0,76-0,94, 0,95 ДИ). SHAP-анализ подтвердил значимость спектральных признаков в интерпретации модели. Таким образом, предложенный способ является эффективным инструментом для ранней диагностики пародонтита, обладающим высокой чувствительностью, специфичностью и интерпретируемостью.

Спектральные изменения при пародонтите, установленные в данной работе, не проявляются при других распространенных стоматологических заболеваниях, например, при кариесе зубов. Ранее были проведены такие исследования [148], которые показали, что при кариесе снижается концентрация ионов $(PO_4)^{3-}$, что было также показано и в работе [149].

Проведена *in vivo*-валидация разработанной модели на 40 спектрах эмали добровольцев (диагноз при измерениях был неизвестен): при пороге $p > 0,5$ модель классифицировала 35 из них (88 %) как «пародонтит», причём 28 спектров (70 %) получили очень высокую вероятность ($p > 0,8$), 7 спектров (17,5 %) – пограничные оценки ($0,3 < p \leq 0,7$) и лишь 2 спектра (5 %) – здоровые ($p \leq 0,3$). Распределение оценок сосредоточено в области высоких значений (около 0,9), что свидетельствует о высокой уверенности алгоритма и редкости «пограничных» случаев

Диагностирование спектральных изменений в эмали зубов, а также разработанный способ верифицирования эмали с диагнозом пародонтит позволит идентифицировать пациентов с группой риска и проводить коррекцию их лечения с помощью гидроксиапатита. Полученные результаты являются предпосылкой для создания экспресс-прибора неинвазивной (*in vivo*) оценки пародонтита по изменению спектральных показателей эмали зубов.

2. В рамках настоящего исследования проведена спектроскопическая оценка состояния костной ткани при пародонтите. Всего было проанализировано 72 спектра (46 – пародонтит, 26 – норма). Выявлены существенные различия в

спектрах КР, проявляющиеся в повышенной интенсивности линий ~ 1318 и ~ 1744 см^{-1} при пародонтите, что может быть связано с изменениями органического матрикса и присутствием бактериальной ДНК, тогда как в контрольной группе более выражены линии, соответствующие нормальному минеральному и белковому составу (~ 852 , ~ 1039 , ~ 1067 , ~ 1244 , ~ 1444 , ~ 1663 см^{-1}).

Была выполнена декомпозиция спектров с высокой степенью точности ($R^2 = 99,82\%$, $\chi^2 = 2 \times 10^{-6}$, AIC = -16262 , BIC = -14267), проведён многоэтапный отбор признаков, и сформирован информативный поднабор из 5 спектральных параметров. Статистический анализ подтвердил наличие достоверных различий между группами (пермутационный тест: $p = 0,00081$; мощность = $0,99$), что доказывает релевантность выбранных признаков.

Модель логистической регрессии (ROC AUC = $0,88$ ($0,78$ - $0,96$, $0,95$ ДИ), F1-мера = $0,72$ ($0,55$ - $0,86$, $0,95$ ДИ), чувствительность = $0,71$ ($0,59$ - $0,82$, $0,95$ ДИ), специфичность = $0,71$ ($0,59$ - $0,82$, $0,95$ ДИ)) успешно решает задачу классификации состояний костной ткани. SHAP-анализ показал, что признаки на 1318 и 1744 см^{-1} являются ключевыми маркерами патологического процесса. Таким образом, в результате исследования разработан надёжный оптический критерий, позволяющий с высокой степенью достоверности идентифицировать изменения костной ткани, связанные с пародонтитом, что делает способ перспективным для клинической диагностики.

3. Проведена оценка спектральных характеристик твёрдых тканей зубов (эмаль, дентин, цемент) до и через шесть недель после проведения закрытого кюретажа. Несмотря на наблюдаемые спектральные изменения, выражающиеся в усилении интенсивности линий гидроксиапатита (~ 432 , ~ 589 , ~ 957 см^{-1}) и карбонат-ионов (~ 1068 см^{-1}), статистический анализ (пермутационный тест: $p > 0,01$) и МГК показали отсутствие достоверных различий между группами. Это позволяет сделать вывод об отсутствии статистически значимого эффекта процедуры лечения на спектральный состав тканей зуба в исследуемом временном интервале.

Вероятно, однократное вмешательство и шестинедельный период наблюдения оказались недостаточными для проявления устойчивых биохимических изменений, требующих более длительного контроля в динамике. И вероятно процедура обеспечивала сходные условия для образования биопленки и зубного налета, репарации тканей и нового прикрепления к поверхности корня.

ЗАКЛЮЧЕНИЕ

В настоящей диссертационной работе предложен и экспериментально подтверждён способ анализа спектров комбинационного рассеяния для диагностики признаков пародонтита и оценки качества дентинных материалов. Доказана возможность неинвазивного контроля состояния твёрдых тканей зубов и костной ткани.

1. Разработан способ анализа спектров комбинационного рассеяния биоматериалов, включающий этапы интерполяции, нормализации, коррекции базовой линии и сглаживания спектральных данных, а также оригинальный итеративный алгоритм декомпозиции спектров. Данный подход позволяет выделять элементарные спектральные компоненты даже при высоком уровне взаимного перекрытия линий до 75 % и шумовых помех, что существенно снижает ошибку интерпретации исходных данных.

Алгоритм декомпозиции отличается от других известных применением кластеризации и генерацией итоговой модели на основе множества спектров. Алгоритм позволяет учитывать сдвиг положения линий КР биологических материалов и других параметров, работу с асимметричными линиями. Позволяет корректно выявлять линии за счет определения итоговой модели на основе композиции множества отдельных моделей.

Алгоритм обеспечивает точность определения позиций линий не хуже $0,1 \text{ см}^{-1}$, относительную погрешность амплитуд и ширин $\leq 0,3 \%$ и детерминацию модели $R^2 \geq 0,995$.

2. Предложен способ оценки деминерализации дентинных материалов по спектрам комбинационного рассеяния, основанный на извлечении информативных признаков из декомпозированных спектральных линий в диапазоне $630\text{--}1450 \text{ см}^{-1}$ с применением логистической регрессии для вероятностной классификации образцов, позволяющей с точностью ROC AUC 1,0 (0,99–1,0, 95 % ДИ) определять факт деминерализации.

3. Предложен способ экспресс-диагностики признаков пародонтита по спектрам эмали и костной ткани с применением логистической регрессии. Способ позволяет проводить диагностику на уровне ROC AUC 0,94 (0,89-0,98, 0,95 ДИ) для эмали и 0,88 (0,78-0,96, 0,95 ДИ) для костной ткани. Способ демонстрирует чувствительность и специфичность на уровне 0,86 (0,76-0,94, 0,95 ДИ) и 0,86 (0,76-0,94, 0,95 ДИ).

Способ рекомендован к применению для диагностики пародонтита у пациентов возрастной группы 35-70 лет.

Перспективы дальнейшей разработки по теме диссертации предполагают интеграцию способа в портативные диагностические инструменты, работающие в режиме реального времени, что повысит их клиническую применимость для стоматологической и медицинской диагностики. Существует значительный потенциал для внедрения алгоритмов машинного обучения для повышения точности и эффективности спектрального анализа, что позволяет более точно выявлять патологические изменения в тканях.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

- ДИ – Доверительный интервал
- МГК – метод главных компонент (англ. Principal Component Analysis, PCA)
- МНК – метод наименьших квадратов (англ. Least squares approximation)
- КР – комбинационное рассеяние
- КТ – компьютерная томография
- ОКТ – Оптическая когерентная томография
- РЭМ – растровый электронный микроскоп
- ЯМР - ядерный магнитный резонанс
- AIC – критерий информационной теории Акаике (англ. Akaike Information Criterion)
- ANOVA - дисперсионный анализ (англ. Analysis of Variance)
- BIC – байесовский информационный критерий (англ. Bayesian Information Criterion)
- CNN – сверточная нейронная сеть (англ. Convolutional Neural Network)
- CTGAN - генеративно-состязательная сеть для работы с табличными данными (англ. Conditional Tabular GAN)
- FTIR - Фурье-ИК-спектроскопия (англ. Fourier Transform Infrared Spectroscopy)
- EMSC – расширенная мультипликативная коррекция сигнала (англ. Extended Multiplicative Signal Correction)
- HWHM – половина максимальной ширины на полувывсоте (англ. Half Width at Half Maximum)
- I-ModPoly – Улучшенный модифицированный метод многополиномиального аппроксимирования (англ. Improved Modified Multi-Polynomial Fitting)
- LOOCV - Поэлементная кросс-валидация (англ. Leave-one-out cross-validation)
- MAE – средняя абсолютная ошибка (англ. Mean Absolute Error)

MAPE – средняя абсолютная процентная ошибка (англ. Mean Absolute Percentage Error)

MLE-SG – Фильтр максимального правдоподобия Савицкого-Голея (англ. Maximum Likelihood Estimation Savitzky-Golay)

PLS-DA – Дискриминантный анализ методом частичных наименьших квадратов (англ. Partial Least Squares Discriminant Analysis)

PS-OCT – Поляризационно-чувствительная оптическая когерентная томография

QLF – Количественная светоиндуцированная флюоресценция

R^2 – коэффициент детерминации (англ. R-squared)

RFE – рекурсивное удаление признаков (англ. Recursive feature elimination)

RMSE – среднеквадратичная ошибка (англ. Root Mean Square Error)

ROC-AUC – площадь под кривой ошибок классификации (англ. Receiver Operating Characteristic - Area Under Curve)

SS-OCT – Swept-source оптическая когерентная томография

TPR – доля истинно положительных предсказаний (англ. True Positive Rate)

VMD – вариационная модовая декомпозиция (англ. Variational Mode Decomposition)

СЛОВАРЬ ТЕРМИНОВ

Аллотрансплантаты и ксеноимплантаты: донорские ткани, полученные от индивидов того же вида (аллотрансплантаты) или другого вида (ксеноимплантаты), подверженные специальной обработке для уменьшения иммуногенности при трансплантации.

Биоматериалы: материалы, используемые для замещения, восстановления или регенерации повреждённых тканей организма, причем они могут иметь как натуральное, так и синтетическое происхождение.

Брефоматрикс: деминерализованный костный матрикс, который получают путём обработки костной ткани.

Вариационная модовая декомпозиция (Variational Mode Decomposition VMD): современный алгоритм декомпозиции сложных сигналов на несколько компонент с различными частотными характеристиками, обеспечивающая эффективное разделение сигналов для повышения точности их анализа.

Гидроксиапатит: минеральное соединение, основная составляющая костной и зубной тканей, с химической формулой $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$. Он придает костям и зубам их прочность и твердость.

Гингивит: воспалительный процесс в десневых тканях, характеризующийся отёком, покраснением и кровоточивостью, который может выступать как предвестник более тяжёлых заболеваний пародонта.

Деминерализация: физиологический или патологический процесс утраты минералов (преимущественно кальция и фосфатов) из зубной или костной ткани, что приводит к ухудшению её механических свойств и структурным изменениям.

Декальцинирование: технологический процесс удаления минерального компонента, преимущественно кальция, из биоматериала (например, дентина) с целью повышения его биоинтегративных свойств, при условии сохранения органических фракций, необходимых для регенерации тканей.

Дентин: органоминеральная структура, составляющая основную массу зуба, расположенная под эмалью и обеспечивающая зубу механическую прочность и амортизационные свойства в условиях жевательного давления.

Депульпирование: это эндодонтическая процедура, заключающаяся в полном удалении пульпы зуба (нервно-сосудистой ткани), которая поражена воспалительным процессом, инфицирована или перешла в состояние некроза. Эта операция является первым этапом лечения корневых каналов: после депульпирования производится формирование и обработка каналов, их пломбирование и герметизация для предотвращения повторного инфицирования и сохранения зуба.

Имплантаты: биосовместимые искусственные конструкции, предназначенные для замены утраченных зубов или восстановления костной ткани посредством остеоинтеграции и обеспечения долговременной функциональной поддержки.

Интерполяция: математический метод, позволяющий оценить значения функции между заданными точками измерений, что используется при обработке спектральных данных для унификации шага дискретизации.

Критерии информационной теории (AIC и BIC): статистические показатели, позволяющие оценить качество и оптимальность моделирования данных с учётом сложности модели, где AIC (критерий Акаике) и BIC (байесовский информационный критерий) штрафуют избыток параметров.

Костно-пластические материалы: материалы, применяемые для восстановления или замещения костной ткани, включающие как живые (аллотрансплантаты), так и синтетические компоненты, обеспечивающие необходимую биосовместимость и механическую прочность при регенеративных процессах.

Количественная светоиндуцированная флюоресценция (QLF): оптический метод диагностики, основанный на измерении интенсивности флюоресценции зубной эмали, позволяющий обнаруживать начальные кариозные изменения с

высокой чувствительностью и специфичностью, особенно в сочетании с алгоритмами машинного обучения.

ЛИОПЛАСТ: метод получения деминерализованного костного или зубного материала, применяемого в стоматологии и челюстно-лицевой хирургии в качестве остеопластического трансплантата. Этапы получения деминерализованного дентина: ультразвуковая очистка, деминерализация раствором соляной кислоты, промывание и нейтрализация, стерилизация с использованием гамма-облучения, сублимационное высушивание.

Логистическая регрессия: статистический метод машинного обучения бинарной классификации, используемый для моделирования вероятности принадлежности наблюдения к одной из двух категорий на основе набора независимых переменных, широко применяемый в диагностических алгоритмах.

MAE, MAPE, RMSE: метрики качества прогностических моделей, где MAE (средняя абсолютная ошибка) характеризует среднее абсолютное отклонение, MAPE (средняя абсолютная процентная ошибка) отражает относительную величину ошибок, а RMSE (корень из среднеквадратичной ошибки) чувствителен к крупным отклонениям в моделировании.

MLE-SG (Maximum Likelihood Estimation Savitzky-Golay): метод сглаживания спектральных данных с использованием фильтра Савицкого-Голея, где параметры оптимизируются посредством метода максимального правдоподобия для повышения точности аппроксимации и подавления шума.

Метод главных компонент (PCA, МГК): статистический алгоритм сокращения размерности данных, при которой исходные переменные преобразуются в новый набор некоррелированных компонентов, сохраняющих максимальную дисперсию, что упрощает последующий анализ сложных многомерных наборов данных.

Метод наименьших квадратов (МНК): оптимизационный алгоритм, минимизирующий сумму квадратов отклонений между экспериментальными данными и предсказанной моделью, что позволяет получить наиболее точные параметры аппроксимации.

Неинвазивность: характеристика диагностических или лечебных методов, при которых не нарушается целостность организма, то есть не происходит проникновения инструментов или веществ внутрь тканей или полостей тела.

Нормализация: процесс обработки данных, в ходе которого исходные значения преобразуются в стандартное пространство, что обеспечивает корректное сравнение и дальнейший анализ разнородных спектральных измерений.

Оптическая когерентная томография (ОКТ) и PS-ОСТ: неразрушающий метод получения высокоразрешённых изображений микроструктур тканей посредством анализа когерентно отражённого света; модификация PS-ОСТ учитывает поляризационные свойства для выявления дополнительных структурных особенностей.

Оптическое поляриметрическое обнаружение: метод исследования, основанный на анализе изменений параметров поляризации света после прохождения через биологическую ткань, что позволяет выявлять микроструктурные аномалии и оценивать оптические свойства тканей.

Пародонтит: воспалительное заболевание тканей пародонта, характеризующееся прогрессирующим разрушением нормальной структуры альвеолярного отростка верхних челюстей или(и) альвеолярной части нижней челюсти. Пародонтит распространён достаточно широко, как и другие заболевания пародонта.

Половина максимальной ширины на полувысоте (HWHM): параметр спектрального анализа, определяемый как ширина пика на уровне 50% от его максимальной интенсивности, используемый для характеристики разрешающей способности спектральных систем.

Реминерализация: биохимический процесс осаждения минеральных компонентов (главным образом кальция и фосфатов) в зубной или костной ткани, способствующий восстановлению их структурной целостности после деминерализации.

Рекурсивное удаление признаков (RFE): итеративный алгоритм отбора наиболее информативных переменных для построения прогностических моделей,

позволяющий повысить их точность и снизить переобучение за счёт устранения незначимых признаков.

ROC-AUC (Receiver Operating Characteristic: Area Under Curve) – показатель эффективности классификационных моделей, представляющий площадь под кривой ROC, где оптимальное соотношение чувствительности и специфичности отражает высокую точность диагностического алгоритма.

Спектроскопия комбинационного рассеяния (КР) или Рамановская спектроскопия: оптический метод исследования, использующий анализ спектров света, полученных в результате комбинационного (рассеянного с участием нескольких колебательных мод) взаимодействия с веществом, что позволяет получать детальную информацию о химическом составе и структурных свойствах биоматериалов.

Фильтр Савицкого-Голея: алгоритмический метод сглаживания спектральных данных, позволяющий уменьшить влияние высокочастотного шума.

Ювенильный дентин: тип зубной ткани, который формируется в молочных (детских) зубах и характеризуется менее высокой степенью минерализации, более высоким содержанием органических компонентов (например, коллагена) и отличительной микроструктурой по сравнению с дентином постоянных зубов.

СПИСОК ЛИТЕРАТУРЫ

1. Епифанова Ю.В., Аванесов А.М. Оценка уровней биоаминов крови у больных хроническим пародонтитом // Электронный научно-образовательный вестник Здоровье и образование в XXI веке. 2012. Vol. 7. No. 14. pp. 114-115.
2. Шашмурина В.Р., Девликанова Л.И., Чумаченко Е.Н. Биомеханические особенности депульпации зубов при заболеваниях пародонта // Российский стоматологический журнал. 2017. Vol. 21. No. 2. pp. 64-67.
3. Papapanou P.N., Susin C. Periodontitis epidemiology: is periodontitis under-recognized, over-diagnosed, or both? // Periodontology. 2017. Vol. 1. No. 75. pp. 45-51.
4. Lindskog S., Blomlof L. Cementum hypoplasia in teeth affected by juvenile periodontitis // J Clin Periodontol. 1983. Vol. 10. pp. 443–451.
5. Jarupat J., Rapeeporn C., Choosak N. Artificial intelligence-powered innovations in periodontal diagnosis: a new era in dental healthcare // Frontiers in Medical Technology. 2025. Vol. 6.
6. Park E.Y., Jeong S., Kang S., Cho J., Cho J.Y., Kim E.K. Tooth caries classification with quantitative light-induced fluorescence (QLF) images using convolutional neural network for permanent teeth in vivo // BMC Oral Health. 2023. Vol. 23. No. 981.
7. Fujii S., Sato S., Fukuda K., Okinaga T., Ariyoshi W., Usui M., Nakashima K., Nishihara T., Takenaka S. Diagnosis of Periodontal Disease from Saliva Samples Using Fourier Transform Infrared Microscopy Coupled with Partial Least Squares Discriminant Analysis // Analytical Sciences. Февраль 2016. Vol. 32. No. 2. pp. 225-231.
8. Silva S.M.S.D., Ferreira C.L., Rizzato J.M.B., Toledo D.S.G., Furukawa M., Rovai E.S., Nogueira M.S. Infrared spectroscopy for fast screening of diabetes and periodontitis // Photodiagnosis Photodyn Ther. Апрель 2024. Vol. 46. P. 104106.
9. Zhang J., Liu Y., Li H., Cao S., Li X., Yin H. Discrimination of periodontal pathogens using Raman spectroscopy combined with machine learning algorithms // Journal of Innovative Optical Health Sciences. 2022. Vol. 15. No. 3.

10. Matsuzaki K., Shimada Y., Shinno Y., Ono S., Yamaji K., Ohara N., Sadr A., Sumi Y., Tagami J., Yoshiyama M. Assessment of Demineralization Inhibition Effects of Dentin Desensitizers Using Swept-Source Optical Coherence Tomography // *Materials*. 2021. Vol. 14. No. 8. P. 1876.
11. Manesh S.K., Darling C.L., Fried D. Nondestructive assessment of dentin demineralization using polarization-sensitive optical coherence tomography after exposure to fluoride and laser irradiation // *J Biomed Mater Res B Appl Biomater*. 2009. Vol. 90. No. 2. pp. 802-812.
12. Denis M., Wan L., Fatemi M., Alizad A. Ultrasound Characterization of Bone Demineralization Using a Support Vector Machine // *Ultrasound in Medicine and Biology*. Март 2018. Vol. 44. No. 3. pp. 714-725.
13. Ломакин М.В. Новая система стоматологических остеоинтегрируемых имплантатов // Автореф. дис. д-ра мед. наук. М.: Московский государственный медицинский стоматологический университет. 2001. P. 49.
14. Кирилова И.А., Садовой М.А., Подорожная В.Т. Сравнительная характеристика материалов для кост- // *Хирургия позвоночника: Новосибирский НИИ травматологии и*. 2012. pp. 72–83.
15. Misch C.E., Dietsch F. Bone-grafting materials in implant dentistry // *Implant Dent*. 1993. Vol. 2. No. 3. pp. 158–167.
16. Murata M., Akazawa T., Mitsugi M., Um I.W., Kim K.W., Ki Y.K. Human dentinas novel biomaterial // *Biomaterials – Physics and Chemistry, InTech*. 2011. pp. 127–137.
17. Usenik P., Bürmen M., Vrtovec T., Fidler A. Automated classification and visualization of healthy and pathological dental tissues based on near-infrared hyperspectral imaging // *Proceedings of SPIE - The International Society for Optical Engineering*. 2011. Vol. 7963.
18. Tramini P., Pélissier B. A Raman Spectroscopic Investigation of Dentin and Enamel Structures Modified by Lactic Acid // *Caries Res*. 2000. pp. 33–240.

19. Camerlingo C., d'Apuzzo F., Grassia V., Perillo L., Lepore M. Micro-Raman spectroscopy for monitoring changes in periodontal ligaments and gingival crevicular fluid // *Sensors (Basel)*. 2014. Vol. 14. No. 12. pp. 22552–22563.
20. Marin E. Raman spectroscopy for early detection and monitoring of dentin demineralization. // *Dent Mater*. 2020.
21. Marin E, Hiraishi N., Honma T., Boschetto F., Zanooco M., Zhu W., Adachi T., Kanamura N., Yamamoto T., Pezzotti G. Raman spectroscopy for early detection and monitoring of dentin demineralization // *Dent Mater*. 2020. Vol. 36. No. 12. pp. 1635-1644.
22. Zhan Z., Zhang X., Guo W., Xie S. Assessment of composition changes in demineralization and remineralization of human dentin using Raman spectroscopy // *Photon Lasers Med*. 2013. Vol. 2. No. 1. pp. 45-50.
23. Чупрунов К.О. Разработка метода получения наноструктурных сферических порошковых материалов на основе гидроксилapatита с регулируемым фазовым составом и показателям дисперсности. Москва: Диссертация на соискание ученой степени кандидата технических наук, 2020. 137 pp.
24. Савельев В.И. Аллотрансплантация формализированной костной ткани в травматологии и ортопедии. СПб: МОРСАРАВ, 2001. 208 pp.
25. Ладонин С.В. Применение аллогенных деминерализованных костных имплантатов при лечении хронического остеомиелита в эксперименте // *Актуальные вопросы тканевой и клеточной трансплантологии: матер. III Всерос. симп. – М. 2007. Р. 27.*
26. Samsonov V.E., Volova L.T., Vasiliev M.V. The use of demineralized bone matrix in the treatment of peri-root cysts and cystogranulomas of the jaws on an outpatient basis // *Kazan medical journal*. 1993. Vol. 74. No. 4. pp. 294-296.
27. Орехова Л.Ю. Заболевания пародонта. Поли Медиа Пресс, 2004. 432 pp.
28. Орехова Л.Ю. Проблемы стоматологического здоровья у лиц молодого возраста (обзор литературы) // *Пародонтология*. 2014. Vol. 2. pp. 3-5.

29. Улитовский С.Б., Алексеева Е.С., Васянина А.А. Проблемы пародонтологии и пути их решения // Пародонтология. 2015. Vol. 3. No. 76. pp. 33-36.
30. Янушевич О.О., Дмитриева Л.А., Ревазова З.Э. Пародонтит XXI век. Москва: МГМСУ, 2012. 366 pp.
31. Егорова Е.Н. Биохимия полости рта. Тверь: Твер. гос. мед. универ., 2018. 155 pp.
32. Berezina N.V., Krivonos S.M. Modern methods of increasing the efficiency of chronic generalized periodontitis treatment. Minsk: BSMU, 2019. 187 pp.
33. Artyushevitch A.S. Periodontal Diseases. Moscow: MedLit, 2006. 328 pp.
34. Shuttle M., Faddy M.J., Cullinan M.P. The clinical course of chronic periodontitis: V. Predictive factors in periodontal disease. // J. Clin. Periodontol. 2009. Vol. 36. pp. 365–371.
35. Серов Л.М., Серова Е.Л., Серов А.Л. Periodontitis: local focus serious problems (literature review) // Parodontologiya. 2014. Vol. 3. No. 19. pp. 3-6.
36. Егорова Е.Н. Биохимия полости рта, учебное пособие для Б 638 студентов стомат. фак. Тверь: Твер. гос. мед. универ., 2018. 155 pp.
37. Н. А. Дорофейчик-Дрыгина Л.Б.Д. Снижение минеральной плотности костной ткани и состояние зубочелюстной системы: клинические параллели // Medical alphabet. 2018. Vol. 2. No. 31. pp. 50-53.
38. Хайдар Д.А., Кульченко А.Г. Гипоксия тканей пародонта при хроническом пародонтите // Специальный выпуск: Материалы XIX международного конгресса “Здоровье и образование в XXI веке”. 2017.
39. Сухинина А.В. Методы оптической спектроскопии для диагностики стоматологических заболеваний. Москва: Автореф. дис. Канд. Физико-математических наук, 2014. 24 pp.
40. Gonchukov S., Sukhinina A., Bakhmutov D., Minaeva S. Raman spectroscopy of saliva as a perspective method for periodontitis diagnostics // Laser Physics Letters. 2012. Vol. 9. No. 1. pp. 73–77.

41. Gatin E., Nagy P., Paun I., Dubok O., Bucur V.W. Raman Spectroscopy: Application in Periodontal and Oral Regenerative Surgery for Bone Evaluation // IRBM. 2019. pp. 279–285.
42. Цимбалистов А.В., Пихур О.Л., Дубова М.А., Садиков Р.А., Франк-Каменецкая О.В., Голованова О.А., Бельская Л.В. Морфология и состав твердых тканей зубов и зубных отложений при генерализованном пародонтите // Вестник Санкт-Петербургского университета. 2006. Vol. 11. No. 1. pp. 128-135.
43. Ремизов С.М. Определение микротвердости для сравнительной оценки зубной ткани здоровых и больных зубов человека // Стоматология. 1965. Vol. 3. pp. 33-37.
44. Ramakrishnaiah R., Rehman G., Basavarajappa S., Khuraif A. Applications of Raman Spectroscopy in Dentistry: Analysis of Tooth Structure // Applied Spectroscopy Reviews. 2014. Vol. 50. No. 4. pp. 332–350.
45. Hsiao T.Y., Lee S.Y., Sun C.W. Optical Polarimetric Detection for Dental Hard Tissue Diseases Characterization // Sensors. 2019. Vol. 19. No. 22.
46. Deng K., Pelekos G., Jin L., Tonetti M.S. Diagnostic accuracy of self-reported measures of periodontal disease: A clinical validation study using the 2017 case definitions // J Clin Periodontol. Август 2021. Vol. 48. No. 8. pp. 1037-1050.
47. Kugahara T., Shosenji Y., Ohashi K. Screening for periodontitis in pregnant women with salivary enzymes // J Obstet Gynaecol Res. Февраль 2008. Vol. 34. No. 1. pp. 40-46.
48. Turton M.S., Henkel R.R., Africa W.J. A simple point of care test can indicate the need for periodontal therapy to reduce the risk for adverse pregnancy outcomes in mothers attending antenatal clinics // Biomarkers. Декабрь 2017. Vol. 22. No. 8. pp. 740-746.
49. Минаева С.А., Михайловский А.А., Бухарова Т.Б., Антонов Е.Н., Гольдштейн Д.В., Попов В.К., Волков А.В. Морфологическое исследование твердых тканей лицевого скелетас помощью спектроскопии комбинационного рассеяния // Российская стоматология. 2015. Vol. 1. pp. 3-10.

50. Otel L. Overall Review on Recent Applications of Raman Spectroscopy Technique in Dentistry // *Quantum Beam Sci.* 2023. Vol. 7. No. 1. P. 5.
51. Zhang Y., Ren L., Wang Q., Wen Z., Liu C., Ding Y. Raman Spectroscopy: A Potential Diagnostic Tool for Oral Diseases // *Frontiers in Cellular and Infection Microbiology.* 2022. Vol. 12.
52. Mandra Y.V., Ivashov A.S., Votjakov S.L., Kiseleva D.V. Possibilities of Raman microspectrometry imaging for structural investigation of human enamel and dentin // *Eksperimentalnaya Klin. Stomatol.* 2011. Vol. 1. pp. 24–28.
53. Ko A.C., Choo-Smith L.P., Hewko M., Sowa M.G., Dong C.C., Cleghorn B. Detection of early dental caries using polarized Raman spectroscopy // *Opt. Express.* 2006. Vol. 14. pp. 203–215.
54. Minaeva S.A., Mikhaylovskiy A.A., Bukharova T.B., Antonov E.N., Goldshteyn D.V., Popov V.K., Volkov A.V. Morphological study of hard tissues of facial skeleton using the Raman spectroscopy // *Russ. Stomatol.* 2005. Vol. 1. pp. 3–10.
55. Korepanov V. Iterative mollifier-based despiking of the single-scan Raman spectra // *Journal of Raman Spectroscopy.* 2022. Vol. 2. pp. 256-259.
56. Barnes R.J., Dhanoa M.S., Lister S.J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra // *Applied Spectroscopy.* 1989. Vol. 5.
57. Afseth N.K., Kohler A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial // *Chemometrics and Intelligent Laboratory Systems.* 2012. pp. 92–99.
58. Lim J.S. *Two-Dimensional Signal and Image Processing* 1990. P. p. 548.
59. Eilers P.H.C. A Perfect Smoother // *Analytical Chemistry.* 2003. Vol. 14.
60. Savitzky A., Golay M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures // *Analytical Chemistry.* 1964. Vol. 8. pp. 1627–39.
61. Blackman R.B., Tukey J.W. *The measurement of power spectra* // *Dover Publications.* 1958.
62. Kanasevich E.R. *Time Sequence Analysis in Geophysics* // *The University of Alberta Press.* 1975. pp. 106-108.

63. Huang N.E., Shen Z., Long S.R. The Empirical Mode Decomposition and The Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis // Proc. R. Soc. A. 1998. Vol. 1971. pp. 903–995.
64. Barton S.J., Ward T.E., Hennesly B.M. Algorithm for optimal denoising of Raman spectra // Analytical Methods. 2018. Vol. 30. pp. 3759-3769.
65. Zhao J., Lui H., McLean D., Zeng H. Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy // Applied spectroscopy. 2007. pp. 1225-32.
66. Lieber C.A., Mahadevan-Jansen A. Automated Method for Subtraction of Fluorescence from Biological Raman Spectra // Applied spectroscopy. 2003. pp. 1363-7.
67. Komsta Ł. Comparison of Several Methods of Chromatographic Baseline Removal with a New Approach Based on Quantile Regression. 2011. No. 73. pp. 721-731.
68. Bertinetto C. Automatic Baseline Recognition for the Correction of Large Sets of Spectra Using Continuous Wavelet Transform and Iterative Fitting // Applied Spectroscopy. 2014. Vol. 68. No. 2. pp. 155-164.
69. Zhi-Min Z., Shan C., Yi-Zeng L. Baseline correction using adaptive iteratively reweighted penalized least squares // Analyst. 2010. Vol. 135. pp. 1138–1146.
70. Schnabel S.K., Eilers P.H.C. Simultaneous estimation of quantile curves using quantile sheets. // AStA Advances in Statistical Analysis. pp. 77-87.
71. Liu J, Sun J., Huang X., Liu B. Goldindex: A Novel Algorithm for Raman Spectrum Baseline Correction // Appl Spectrosc. 2015. Vol. 69. No. 7. pp. 834-42.
72. Eilers P.H.C., Boelens H. Baseline correction with asymmetric least squares smoothing // Leiden University Medical Centre Report. 2005.
73. He S. Baseline correction for Raman spectra using an improved asymmetric least squares method // Analytical Methods. 2014. Vol. 12. pp. 4402-4407.
74. Baek S.J., Park A., Ahn Y.J., Choo J. Baseline correction using asymmetrically reweighted penalized least squares smoothing // The Analyst. 2015. Vol. 1. pp. 250-7.

75. Xu D., Liu S., Cai Y., Yang C. Baseline correction method based on doubly reweighted penalized least squares // *Appl. Opt.* 2019. pp. 3913-3920.
76. Ye J. Baseline correction method based on improved asymmetrically reweighted penalized least squares for Raman spectrum // *Applied Optics*. 2020. Vol. 59. pp. 10933-10943.
77. Zhang F. Baseline correction for infrared spectra using adaptive smoothness parameter penalized least squares method // *Spectroscopy Letters*. 2020. Vol. 53. No. 3. pp. 222-233.
78. Oller-Moreno S. Adaptive Asymmetric Least Squares baseline estimation for analytical instruments // *IEEE 11th International Multi-Conference on Systems, Signals, and Devices*. 2014. pp. 1-5.
79. Korepanov V. Asymmetric least-squares baseline algorithm with peak screening for automatic processing of the Raman spectra // *Journal of Raman Spectroscopy*. 2020. Vol. 51. No. 10. pp. 2061-2065.
80. Li Z. Morphological weighted penalized least squares for background correction // *Analyst*. 2013. Vol. 138. pp. 4483-4492.
81. Perez-Pueyo R. Morphology-Based Automated Baseline Removal for Raman Spectra of Artistic Pigments // *Applied Spectroscopy*. 2010. Vol. 64. pp. 595-600.
82. Dai L. An Automated Baseline Correction Method Based on Iterative Morphological Operations // *Applied Spectroscopy*. 2018. Vol. 72. No. 5. pp. 731-739.
83. Koch M. Iterative morphological and mollifier-based baseline correction for Raman spectra. // *J Raman Spectroscopy*. 2017. Vol. 48. No. 2. pp. 336-342.
84. Chen H. An Adaptive and Fully Automated Baseline Correction Method for Raman Spectroscopy Based on Morphological Operations and Mollifications // *Applied Spectroscop.* 2019. Vol. 73. No. 3. pp. 284-293.
85. Liland K. Optimal Choice of Baseline Correction for Multivariate Calibration of Spectra // *Applied Spectroscopy*. 2010. Vol. 64. No. 9. pp. 1007-1016.
86. Gonzalez-Vidal J. Automatic morphology-based cubic p-spline fitting methodology for smoothing and baseline-removal of Raman spectra // *Journal of Raman Spectroscopy*. 2017. Vol. 48. No. 6. pp. 878-883.

87. Liu H. Joint Baseline-Correction and Denoising for Raman Spectra // *Applied Spectroscopy*. 2015. Vol. 69. No. 9. pp. 1013-1022.
88. de Rooi J. Mixture models for baseline estimation // *Chemometric and Intelligent Laboratory Systems*. 2012. Vol. 117. pp. 56-60.
89. Han Q. 2018 5th International Conference on Information Science and Control Engineering (ICISCE) // *Iterative Reweighted Quantile Regression Using Augmented Lagrangian Optimization for Baseline Correction*. 2018. pp. 280-284.
90. Liu Y.J. A Concise Iterative Method with Bezier Technique for Baseline Construction // *Analyst*. 2015. Vol. 140. No. 23. pp. 7984-7996.
91. Friedrichs M. A model-free algorithm for the removal of baseline artifacts // *J. Biomolecular NMR*. 1995. Vol. 5. pp. 147-153.
92. Ryan C.G. SNIP, A Statistics-Sensitive Background Treatment for The Quantitative Analysis Of Pixe Spectra In Geoscience Applications // *Nuclear Instruments and Methods in Physics Research B*. 1988. Vol. 934. pp. 396-402.
93. Schulze H. A Small-Window Moving Average-Based Fully Automated Baseline Estimation Method for Raman Spectra // *Applied Spectroscopy*. 2012. Vol. 66. No. 7. pp. 757-764.
94. Wang T. Background Subtraction of Raman Spectra Based on Iterative Polynomial Smoothing // *Applied Spectroscopy*. 2017. Vol. 71. No. 6. pp. 1169-1179.
95. Krishna H. Range-independent background subtraction algorithm for recovery of Raman spectra of biological tissue // *J Raman Spectroscopy*. 2012. Vol. 43. No. 12. pp. 1884-1894.
96. Dietrich W. Fast and Precise Automatic Baseline Correction of One- and Two-Dimensional NMR Spectra // *Journal of Magnetic Resonance*. 1991. Vol. 91. pp. 1-11.
97. Golotvin S. Improved Baseline Recognition and Modeling of FT NMR Spectra // *Journal of Magnetic Resonance*. 2000. Vol. 146. pp. 122-125.
98. Wang K.C. Distribution-Based Classification Method for Baseline Correction of Metabolomic 1D Proton Nuclear Magnetic Resonance Spectra // *Analytical Chemistry*. 2013. Vol. 85. pp. 1231-1239.

99. Johnsen L. An automated method for baseline correction, peak finding and peak grouping in chromatographic data // *Analyst*. 2013. Vol. 138. pp. 3502-3511.
100. Cobas J. A new general-purpose fully automatic baseline-correction procedure for 1D and 2D NMR data // *Journal of Magnetic Resonance*. 2006. Vol. 183. No. 1. pp. 145-151.
101. Zhang X., Ma Y., Bai Y., He P., Tang Y., Lv X. Denoising of Raman Spectra Using a Neural Network Based on Variational Mode Decomposition, Empirical Wavelet Transform, and Encoder-Bidirectional Long Short-Term Memory // *Appl. Sci*. 2023. Vol. 13. No. 21. P. 12046.
102. Bian X., Shi Z., Shao Y., Chu Y., Tan X. Variational Mode Decomposition for Raman Spectral Denoising // *Molecules*. 2023. Vol. 28. No. 17. P. 6406.
103. Dong S., Liu Y., Wu J. An Iterative Curve-Fitting Baseline Correction Method for Raman Spectra Driven by Neural Network // *Appl. Spectrosc*. 2024. Vol. 78. No. 1.
104. Mou-Yan Z., Unbehauen R. A deconvolution method for spectroscopy // *Meas. Sci. Technol*. 1995. Vol. 6. pp. 482–487.
105. Liu H., Liu S., Zhang Z. Adaptive total variation-based spectral deconvolution with the split Bregman method // *Appl. Opt*. 2014. No. 53. pp. 8240–8248.
106. Liu H., Zhang T., Yan L. A MAP-based algorithm for spectroscopic semi-blind deconvolution // *Analyst*. 2012. No. 137. pp. 3862–3873.
107. Marquardt F., Alsmeyer W. Automatic generation of peakshaped models // *Applied spectroscopy*. 2004. No. 58. pp. 986–994.
108. Alsmeyer F., Koss H., Marquardt W. Indirect spectral hard modeling for the analysis of reactive and interacting mixtures // *Applied spectroscopy*. 2004. No. 58. pp. 975–985.
109. Meier R.J. On art and science in curve-fitting vibrational spectra // *Vib. Spectrosc*. 2005. No. 39. pp. 266–269.
110. Ida T., Ando M., Toraya H. Extended pseudo-Voigt function for approximating the Voigt profile // *Journal of Applied Crystallography*. 2000. Vol. 33. No. 6. pp. 1311–1316.

111. Chen Y., Dai L. Automated decomposition algorithm for Raman spectra based on a Voigt line profile model // *Appl. Opt.* 2016. No. 55. pp. 4085-4094.
112. Kupriyanov G.A. Decomposition of Spectral Contour into Gaussian Bands using Gender Genetic Algorithm // *Moscow Univ. Phys.* 2023. Vol. 78. No. 1. pp. 236–242.
113. Zhou Q., Zou Z., Han L. Deep Learning-Based Spectrum Reconstruction Method for Raman Spectroscopy // *Coatings.* 2022. Vol. 12. No. 8. P. 1229.
114. Bian X., Shi Z., Shao Y. Variational Mode Decomposition for Raman Spectral Denoising // *Molecules.* 2023. Vol. 28. P. 6406.
115. Schmid N., Bruderer S., Paruzzo F. Deconvolution of 1D NMR spectra: A deep learning-based approach // *Journal of Magnetic Resonance.* 2023. Vol. 347. P. 107357.
116. Kaneki S., Kouketsu Y., Aoya M. An automatic peak deconvolution code for Raman spectra of carbonaceous material and a revised geothermometer for intermediate- to moderately high-grade metamorphism // *Prog Earth Planet Sci.* 2024. Vol. 11. No. 35.
117. Сизиков В.С., Лавров А.В. Сравнение различных методов разделения непрерывных перекрывающихся спектральных линий // *Оптика и спектроскопия.* 2018. Vol. 124. No. 6. pp. 1-9.
118. Branch M.A., Coleman T.F., Li Y. A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems // *SIAM Journal on Scientific Computing.* 1999. Vol. 21. No. 1. pp. 1-23.
119. Levenberg K. A Method for the Solution of Certain Non-Linear Problems in Least Squares // *Quarterly of Applied Mathematics.* 1944. Vol. 2. No. 2. pp. 164–168.
120. Marquardt D. An Algorithm for Least-Squares Estimation of Nonlinear Parameters // *SIAM Journal on Applied Mathematics.* 1963. Vol. 11. No. 2. pp. 431–441.
121. Nelder J.A., Mead R. A simplex method for function minimization // *Computer Journal.* 1965. Vol. 7. No. 4. pp. 308–313.

122. Zhu C., Byrd R.H., Lu P., Nocedal J. L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization // ACM Transactions on Mathematical Software. 1997. Vol. 23. No. 4. pp. 550–560.
123. Powell M.J.D. An efficient method for finding the minimum of a function of several variables without calculating derivatives // Computer Journal. 1964. Vol. 7. No. 2. pp. 155–162.
124. Hestenes M.R., Stiefel E. Methods of Conjugate Gradients for Solving Linear Systems // Journal of Research of the National Bureau of Standards. 1952. Vol. 49. No. 6. P. 409.
125. Byrd R.H., Lu P., Nocedal J., Zhu C. A Limited Memory Algorithm for Bound Constrained Optimization // SIAM Journal on Scientific Computing. 1995. Vol. 16. No. 5. pp. 1190–1208.
126. Dembo R.S., Eisenstat S.C., Steihaug T. Inexact newton methods // SIAM Journal on Numerical analysis. 1982. Vol. 19. No. 2. pp. 400-408.
127. Ester M., Kriegel H.P., Sander J., Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise // Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. 1996. pp. 226-231.
128. Wolberg J. Data Analysis Using the Method of Least Squares. Springer, 2006.
129. Cheng W.T., Liu M.T., Liu H.N., Lin S.Y. Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma // Microscopy Research and Technique. 2005. Vol. 68. pp. 75–79.
130. Chan J.W., Taylor D.S., Zwerdling T., Lane S.T., Ihara K., Huser T. Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells // Biophysical Journal. 2006. Vol. 90. pp. 648–656.
131. Silveira L., Sathaiyah S., Zangaro R.A. Correlation between nearinfrared Raman spectroscopy and the histopathological analysis of atherosclerosis in human coronary arteries // Lasers in Surgery and Medicine. 2002. Vol. 30. pp. 290–297.
132. Lundberg S.M., Lee S.I. A Unified Approach to Interpreting Model Predictions // Curran Associates. 2017. No. 30.

133. Udaltsova K.A. The structure of intact dentin and enamel of human juvenile teeth // *World Med. Biol.* 2008. Vol. 4. pp. 70–76.
134. Lakshimi R.J. Tissue Raman spectroscopy for the study of radiation damage: Brain irradiation of mice // *Radiation Research.* 2002. Vol. 157. pp. 175–182.
135. Ruiz-Chica A.J., Medina M.A., Sanchez-Jimenez F., Ramirez F.J. Characterization by Raman spectroscopy of conformational changes on guaninecytosine and adenine-thymine oligonucleotides induced by aminoxy analogues of spermidine // *Journal of Raman Spectroscopy.* 2004. No. 35. pp. 93-100.
136. Kateinen E., Elomaa M., Laakkonen U.M., Sippola E., Niemela P., Suhonen J., Jarninen K. Qualification of the amphetamine content in seized street samples by Raman spectroscopy // *Journal of Forensic Science.* 2007. Vol. 52. No. 1. pp. 88–92.
137. Souza R., Jerônimo D., Gouvêa H. Fourier-Transform Raman Spectroscopy Study of the Ovariectomized Rat Model of Osteoporosis // *The Open Bone Journal.* 2010. Vol. 2. pp. 24-31.
138. Таболина Е.С., Бородовицина С.И., Савельева Н.А. Профилактика стоматологических заболеваний. Рязань: ФГБОУВОРяз ГМУ Минздрава России, 2019. 279 pp.
139. Stone N., Kendell C., Shepherd N., Crow P., Barr H. Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers // *Journal of Raman Spectroscopy.* 2002. Vol. 33. pp. 564–573.
140. Cárdenas A.M., Ardila L.J., Vernal R., Melgar-Rodríguez S., Hernández H.G. Biomarkers of Periodontitis and Its Differential DNA Methylation and Gene Expression in Immune Cells: A Systematic Review // *Int. J. Mol. Sci.* 2022. Vol. 23. P. 12042.
141. Konečná B., Gaál Kovalčíková A., Pančíková A., Novák B., Kovařová E., Celec P., Tóthová L. Salivary Extracellular DNA and DNase Activity in Periodontitis // *Applied Sciences.* 2020. Vol. 10. No. 21. P. 7490.
142. Zhang Y., Kang N., Xue F. Evaluation of salivary biomarkers for the diagnosis of periodontitis // *BMC Oral Health.* 2021. Vol. 21. No. 266.

143. Wolf H.F., Rateitschak E.M. Periodontology. Moscow: MED press-Inform, 2014. 548 pp.
144. Gao L., Liu Y., Kim D., Li Y., Hwang G., Naha P.C., Cormode D.P., Koo H. Nanocatalysts promote *Streptococcus mutans* biofilm matrix degradation and enhance bacterial killing to suppress dental caries in vivo // *Biomaterials*. 2016. Vol. 101. pp. 272–284.
145. Lyamin A.V., Botrin E.A., Zhestkov A.V. Medical problems associated with bacterial biofilms // *Clin. Microbiol. Antimicrob. Chemother.* 2012. Vol. 14. pp. 268–275.
146. Ricci G., Aimetti M. Periodontal Diagnosis and Therapy. Moscow: Azbuka, 2015. 739 pp.
147. James S.S., Gary S.Y., Leif B.K. Comparison of cellular cementum in normal and diseased teeth-a scanning electron microscopic study // *J. Endod.* 1981. Vol. 7. pp. 370–378.
148. Timchenko P.E. Use of Raman spectroscopy for diagnosis of disease in dental tissue // *J. Opt. Technol.* 2016. Vol. 83. No. 5. pp. 313-317.
149. Мандра Ю.В., Ивашов А.С., Вотяков С.Л., Киселева В.В. Возможности применения рамановской микроспектроскопии для исследования структурных особенностей твердых тканей зубов человека // *Экспериментально-клиническая стоматология*. 2011. Vol. 1. pp. 24-28.

ПРИЛОЖЕНИЕ А. Акт о внедрении результатов диссертационной работы в
СамГМУ



федеральное государственное бюджетное
образовательное учреждение высшего образования
«Самарский государственный медицинский университет»
Министерства здравоохранения Российской Федерации
(ФГБОУ ВО СамГМУ Минздрава России)

443099, РФ, г. Самара,
ул. Чапаевская, 89
тел.: +7 (846) 374-10-01
тел./факс: +7 (846) 374-10-03

е-mail: info@samsmu.ru
сайт: samsmu.ru
ОГРН 1026301426348
ИНН 6317002858

УТВЕРЖДАЮ

Проректор по научной работе

ФГБОУ ВО СамГМУ Минздрава России

Д.м.н., профессор



И.Л. Давыдкин

2025 г.

10.07.2025

№ 1130/40-23-3345

На № _____ от _____

АКТ

о внедрении результатов

кандидатской диссертационной работы

Фролова Олега Олеговича

Комиссия в составе:

председатель: директор НИИ БиoТех СамГМУ д.м.н., профессор Л.Т. Волова

члены комиссии: главный специалист НИИ БиoТех СамГМУ И.Ф. Нефедова, инженер НИИ БиoТех СамГМУ А.А. Силютин

составили настоящий акт о том, что результаты диссертационной работы Фролова Олега Олеговича «Разработка способа анализа спектров комбинационного рассеяния для применения в стоматологии», представленной на соискание ученой степени кандидата технических наук, использованы в процессе производства дентинных материалов в научно-исследовательском институте биотехнологий (НИИ БиoТех) Самарского государственного медицинского университета (СамГМУ).

Результаты диссертационной работы, используемые в процессе производства дентинных материалов:

Предложен способ оценки степени деминерализации дентинных материалов по спектрам комбинационного рассеяния, основанный на извлечении информативных признаков из декомпозированных спектральных линий в диапазоне 630–1450 см⁻¹ с применением методов машинного обучения для вероятностной классификации образцов, позволяющей с точностью ROC AUC 1,0 (0,99–1,0, 95% CI) определять факт деминерализации.

Заключение: Использование указанных результатов позволяет оптимизировать процесс изготовления биоматериалов и проводить оценку их качества.

Председатель комиссии:
Директор НИИ БиoТех СамГМУ
д.м.н., профессор

Л.Т. Волова

Члены комиссии:
Главный специалист НИИ БиoТех СамГМУ

И.Ф. Нефедова

Инженер НИИ БиoТех СамГМУ

А.А. Силютин

ПРИЛОЖЕНИЕ Б. Акт о внедрении результатов диссертационной работы в
«Центр восстановительной стоматологии»

УТВЕРЖДАЮ

Главный врач стоматологической клиники
«Центр восстановительной стоматологии»

Глубоков Д.Г.

« 15 » _____ 01 _____ 2025 г.

АКТ

о внедрении результатов

кандидатской диссертационной работы
Фролова Олега Олеговича

Настоящий акт подтверждает, что результаты диссертационной работы Фролова Олега Олеговича «Разработка способа анализа спектров комбинационного рассеяния для применения в стоматологии», представленной на соискание ученой степени кандидата технических наук, используются стоматологической клиникой в процессе экспресс-диагностики пародонтита зубов и полезны в терапевтической и хирургической стоматологии.

Заключение: Использование указанных результатов позволяет повысить точность диагностики пародонтита зубов.

Главный врач стоматологической клиники
«Центр восстановительной
стоматологии»



Глубоков Д.Г.